

16869N-104800US

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2 0 0 3 年 9 月 2 6 日

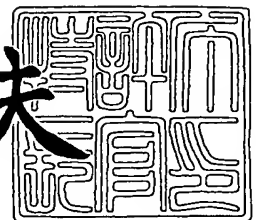
出 願 番 号
Application Number: 特 願 2 0 0 3 - 3 3 5 4 6 5
[ST. 10/C]: [J P 2 0 0 3 - 3 3 5 4 6 5]

出 願 人
Applicant(s): 株式会社日立製作所

2 0 0 3 年 1 0 月 9 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 3 - 3 0 8 3 3 7 3

【書類名】 特許願
【整理番号】 NT03P0165
【提出日】 平成15年 9月26日
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 11/00
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 八木沢 育哉
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 岡本 岳樹
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 松並 直人
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 福岡 幹夫
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 中野 俊夫
【発明者】
 【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A
 I D システム事業部内
 【氏名】 ▲高▼本 賢一
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム研究所内
 【氏名】 山本 彰
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 100068504
 【弁理士】
 【氏名又は名称】 小川 勝男
 【電話番号】 03-3661-0071
【選任した代理人】
 【識別番号】 100086656
 【弁理士】
 【氏名又は名称】 田中 恭助
 【電話番号】 03-3661-0071

【選任した代理人】

【識別番号】 100094352

【弁理士】

【氏名又は名称】 佐々木 孝

【電話番号】 03-3661-0071

【手数料の表示】

【予納台帳番号】 081423

【納付金額】 21,000円

【提出物件の目録】

【物件名】 特許請求の範囲 1

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

複数のディスク装置を有するアレイ型ディスク装置であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値レベル 1 を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリングを開始し、前記ディスク装置のエラー発生回数が前記規定値レベル 1 よりも大きい規定値レベル 2 を超えた場合に前記スペアディスク装置から読み出し処理を行うことを特徴とするアレイ型ディスク装置。

【請求項 2】

複数のディスク装置を有するアレイ型ディスク装置であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値レベル 1 を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、前記ディスク装置のエラー発生回数が前記規定値レベル 1 よりも大きい規定値レベル 2 を超えた場合に前記ディスク装置の閉塞開始を指示し、該ディスク装置で行っていた処理の前記スペアディスク装置への移行を指示するエラー監視部と、前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラー部と、前記ディスク装置の閉塞と前記移行とを行う閉塞移行部とを備えたことを特徴とするアレイ型ディスク装置。

【請求項 3】

複数のディスク装置を有するアレイ型ディスク装置であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に、該ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、ミラーリングをしていないディスク装置のエラー発生回数がミラーリングをしているディスク装置のエラー発生回数を超えた場合、該スペアディスク装置のミラーリングを解除し、ミラーリングをしていないディスク装置とミラーリングを解除したスペアディスク装置とのミラーリングを開始するよう指示するエラー監視部と、

前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラー部とを備えたことを特徴とするアレイ型ディスク装置。

【請求項 4】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に前記ディスク装置のステータスが仮閉塞状態となるよう指示するエラー監視部と、

ディスクアレイグループを構成するディスク装置が仮閉塞状態となった場合、仮閉塞となったディスク装置のデータをディスクアレイグループを構成する他のディスク装置からスペアディスク装置に復旧するデータ復旧部とを備え、

前記データ復旧部は、データ復旧中にディスクアレイグループを構成する他のディスク装置からの読み出しができない場合、仮閉塞状態のディスク装置からの読み出しを行うこ

とを特徴とするアレイ型ディスク装置。

【請求項 5】

複数のディスク装置を有するアレイ型ディスク装置であって、ディスク装置間のデータ移行に際し、移行元ディスク装置からのデータ読み込みエラー発生回数を記憶し、エラー発生回数が規定値に達するまでは移行元ディスク装置から移行先ディスク装置にデータを読み込み、エラー発生回数が規定値に達した場合、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みに切り替え、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みがエラーになり、データ復旧できない場合に、移行元のディスク装置からのデータ読み込みを行うことを特徴としたアレイ型ディスク装置。

【請求項 6】

請求項 5 に記載のアレイ型ディスク装置であって、前述のディスクアレイグループを構成するデータディスク装置からのデータ読み込みに切り替え後、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みがエラーになり、移行元のデータディスク装置からのデータ読み込みに切り替え、その読み込みが成功した場合、移行元のデータディスク装置から読み込んだデータとディスクアレイグループを構成するデータディスク装置からのデータを用いて、読み込みエラーになったディスクアレイグループを構成する前記データディスク装置のデータを復旧することを特徴としたアレイ型ディスク装置。

【請求項 7】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置であって、前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスクアレイ型ディスク装置は、

前記ディスク装置が形成するディスクアレイグループを 1 単位としてディスク装置のエラー発生状況を監視し、エラー発生回数が規定値を超えた場合、該ディスク装置のデータをスペアディスク装置に移行することを指示するエラー監視部と、

前記ディスクアレイグループの複数の前記ディスク装置のエラー発生回数が、規定値よりも小さく設定した補助規定値に達した場合、上記規定値をより小さな値に動的に変更するエラー回数規定値変更部と、

前記移行指示を受けてデータコピーを行うコピー部とを備えたことを特徴とするアレイ型ディスク装置。

【請求項 8】

複数のディスク装置を有するアレイ型ディスク装置におけるエラー監視制御プログラムであって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置におけるエラー監視制御プログラムは、前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値レベル 1 を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリングを開始し、前記ディスク装置のエラー発生回数が前記規定値レベル 1 よりも大きい規定値レベル 2 を超えた場合に前記スペアディスク装置から読み出し処理を行うプログラムであることを特徴とするアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 9】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置におけるエラー監視制御プログラムであって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置におけるエラー監視制御プログラムは、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規

定値レベル 1 を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリング開始を指示するエラー監視プログラムと、

前記ディスク装置のエラー発生回数が前記規定値レベル 1 よりも大きい規定値レベル 2 を超えた場合に前記ディスク装置の閉塞開始を指示し、該ディスク装置で行っていた処理の前記スペアディスク装置への移行を指示する閉塞監視プログラムと、

前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラープログラムと、前記ディスク装置の閉塞と前記移行とを行う閉塞移行プログラムとを備えたことを特徴とするアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 10】

複数のディスク装置を有するアレイ型ディスク装置におけるエラー監視制御プログラムであって、

前記アレイ型ディスクアレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置におけるエラー監視制御プログラムは、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に、前記ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、ミラーリングをしていないディスク装置のエラー発生回数がミラーリングをしているディスク装置のエラー発生回数を超えた場合、該スペアディスク装置のミラーリングを解除し、ミラーリングをしていないディスク装置とミラーリングを解除したスペアディスク装置とのミラーリングを開始するよう指示するエラー監視プログラムと、

前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラープログラムとを備えたことを特徴とするアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 11】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置におけるエラー監視制御プログラムであって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置におけるエラー監視制御プログラムは、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に前記ディスク装置のステータスが仮閉塞状態となるよう指示するエラー監視プログラムと、

ディスクアレイグループを構成するディスク装置が仮閉塞状態となった場合、仮閉塞となったディスク装置のデータをディスクアレイグループを構成する他のディスク装置からスペアディスク装置に復旧するデータ復旧プログラムとを備え、

前記データ復旧プログラムは、データ復元中にディスクアレイグループを構成する他のディスク装置からの読み出しができない場合、仮閉塞状態のディスク装置からの読み出しを行うことでデータ復旧を行うプログラムであることを特徴とするアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 12】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置におけるエラー監視制御プログラムであって、ディスク装置間のデータ移行に際し、移行元ディスク装置からのリードエラー発生回数するエラー発生回数記憶プログラムと、エラー発生回数が規定値に達するまでは移行元のデータをディスク装置から移行先ディスク装置にデータを読み込み、エラー発生回数が規定値に達した場合、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みに切り替え、ディスクアレイグループを構成するデータディスク装置からのデータリードがエラーになり、データ復旧できない場合のみ移行元のディスク装置からのデータ読み込みを行うデータ監視移行プログラムとを有することを特徴としたアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 13】

請求項 12 に記載のアレイ型ディスク装置におけるエラー監視制御プログラムであって

、前述のディスクアレイグループを構成するデータディスク装置からのデータ読み込みに切り替え後、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みがエラーになり、データ復旧できない場合のみ、移行元のディスク装置からのデータ読み込みに切り替えその読み込みに成功した場合、移行元のディスク装置からの読み込んだデータとディスクアレイグループを構成するデータディスク装置からのデータを用いて読み込みエラーになったディスクアレイグループを構成するディスク装置のデータを復旧するデータ復旧プログラムを有することを特徴としたアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 14】

複数のディスク装置を有するアレイ型ディスク装置におけるエラー監視制御プログラムであって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置におけるエラー監視制御プログラムは、

前記ディスク装置が形成するディスクアレイグループを 1 単位としてエラー発生状況を監視し、エラー発生回数が規定値を超えた場合、該ディスク装置のデータをスペアディスク装置に移行することを指示するエラー監視プログラムと、

前記ディスク装置が形成するディスクアレイグループの複数の前記ディスク装置のエラー発生回数が規定値よりも小さく設定した補助規定値に達した場合、上記規定値をより小さな値に動的に変更する規定の条件を超えた場合、ディスクアレイグループを形成する前記ディスク装置のエラー発生回数の規定値を動的に変更するエラー回数規定値変更プログラムと、

前記移行指示を受けてデータコピーを行うコピープログラムとを備えたことを特徴とするアレイ型ディスク装置におけるエラー監視制御プログラム。

【請求項 15】

複数のディスク装置を有するアレイ型ディスク装置のエラー監視制御方法であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置のエラー監視制御方法は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値レベル 1 を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリングを開始するエラー監視方法と、前記ディスク装置のエラー発生回数が前記規定値レベル 1 よりも大きい規定値レベル 2 を超えた場合に前記スペアディスク装置から読み出し処理を行う読み出し方法を含むことを特徴とするアレイ型ディスク装置のエラー監視制御方法。

【請求項 16】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置のエラー監視制御方法であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも 1 台はスペアディスク装置であり、

前記アレイ型ディスク装置のエラー監視制御方法は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値レベル 1 を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、前記ディスク装置のエラー発生回数が前記規定値レベル 1 よりも大きい規定値レベル 2 を超えた場合に前記ディスク装置の閉塞開始を指示し、該ディスク装置で行っていた処理の前記スペアディスク装置への移行を指示するエラー監視方法と、

前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラー方法と、前記ディスク装置の閉塞と前記移行とを行う閉塞移行方法とを含むことを特徴とするアレイ型ディスク装置のエラー監視制御方法。

【請求項 17】

複数のディスク装置を有するアレイ型ディスク装置のエラー監視制御方法であって、
前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも1台はスペアディスク装置であり、

前記アレイ型ディスク装置のエラー監視制御方法は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に、前記ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、ミラーリングをしていないディスク装置のエラー発生回数がミラーリングをしているディスク装置のエラー発生回数を超えた場合、スペアディスク装置のミラーリングを解除し、ミラーリングをしていないディスク装置とミラーリングを解除したスペアディスク装置とのミラーリングを開始するよう指示するエラー監視方法と、

前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラー監視方法とを含むことを特徴とするアレイ型ディスク装置のエラー監視制御方法。

【請求項18】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置のエラー監視制御方法であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも1台はスペアディスク装置であり、

前記アレイ型ディスク装置のエラー監視制御方法は、

前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に前記ディスク装置のステータスが仮閉塞状態となるよう指示するエラー監視方法と、

ディスクアレイグループを構成するディスク装置が仮閉塞状態となった場合、仮閉塞となったディスク装置のデータをディスクアレイグループを構成する他のディスク装置からスペアディスク装置に復旧するデータ復旧方法とを含み、

前記データ復旧方法は、データ復旧中にディスクアレイグループを構成する他のディスク装置からの読み出しができない場合、仮閉塞状態のディスク装置からの読み出しを行う方法であることを特徴とするアレイ型ディスク装置のエラー監視制御方法。

【請求項19】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置のデータ移行方法であって、前記ディスク装置間のデータ移行に際し、移行元ディスク装置からのデータ読み込みエラー発生回数するエラー記憶方法と、エラー発生回数が規定値に達した場合、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みに切り替え、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みがエラーになり、データ復旧できない場合に、移行元のディスク装置からのデータ読み込みを行うデータ監視移行方法とを含むことを特徴としたアレイ型ディスク装置のデータ移行方法。

【請求項20】

請求項19に記載のアレイ型ディスク装置のデータ移行方法であって、前述のディスクアレイグループを構成するデータディスク装置からのデータリード・データ復元方式に切り替え後、冗長度をもつデータディスク装置からのデータリードがエラーになり、データ復元できない場合のみ、移行元のディスクディスク装置からのデータ読み込みに切り替え後、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みがエラーになり、移行元のデータディスク装置からのデータ読み込みに切り替え、その読み込みが成功した場合、移行元のディスクディスク装置からの読み込んだデータとディスクアレイグループを構成するデータディスク装置からのデータを用いて読みエラーになったディスクアレイグループを構成するディスクディスク装置のデータを復旧するデータ復旧方法を含むことを特徴としたアレイ型ディスク装置のデータ移行方法。

【請求項21】

アレイ状に配置した複数のディスク装置を有するアレイ型ディスク装置のエラー監視制御方法であって、

前記アレイ型ディスク装置が有するディスク装置のうち、少なくとも1台はスペアディ

スク装置であり、

前記アレイ型ディスクアレイ型ディスク装置のエラー制御方法は、

前記ディスク装置が形成するディスクアレイグループを 1 単位としてディスク装置のエラー発生状況を監視し、エラー発生回数が規定値を超えた場合、該ディスク装置のデータをスペアディスク装置に移行することを指示するエラー監視方法と、

前記ディスク装置が形成するディスクアレイグループのエラー発生回数が、規定値よりも小さく設定した補助規定値に達した場合、上記規定値をより小さな値に動的に変更するエラー回数規定値変更方法と、

前記移行指示を受けてデータコピーを行うコピー方法とを含むことを特徴とするアレイ型ディスク装置のエラー監視制御方法。

【書類名】 明細書

【発明の名称】 ドライブの 2 重障害を防止するアレイ型ディスク装置、プログラム、及び方法

【技術分野】**【0 0 0 1】**

本発明は主として、コンピュータの外部記憶装置であるディスク装置に関わり、特に、ディスクアレイを構成するアレイ型ディスク装置において複数台のディスク装置が同時に故障を起こすことを予防するための技術に関わり、さらに、冗長性を有するディスクアレイグループを構成するディスク装置間のデータ移行時における、ホスト I / O レスポンス向上及び確実性向上技術に関する。

【背景技術】**【0 0 0 2】**

計算機に接続される記憶装置システム的一种に、アレイ型ディスク装置がある。アレイ型ディスク装置は、R A I D (Redundant Arrays of Inexpensive Disks)とも呼ばれ、アレイ状に配置された複数のディスク装置及びそれらを制御する制御部とを有する記憶装置である。アレイ型ディスク装置では、リード要求（データの読み出し要求）およびライト要求（データの書き込み要求）がディスク装置の並列動作によって高速に処理され、かつデータに冗長性が付加される。アレイ型ディスク装置は、非特許文献 1 に開示されているように、付加される冗長データの種類とその構成により 5 つのレベルに分類されている。

【0 0 0 3】

市場に出回っているアレイ型ディスク装置においては、運用されているディスク装置が故障する場合を想定し、スペアディスク装置をあらかじめ同一アレイ型ディスク装置内に搭載しておくことが一般的である。アレイ型ディスク装置のレイド（R A I D）グループ即ちディスクアレイグループを形成するディスク装置が故障状態になったとアレイ型ディスク装置が判定した場合に、他のディスク装置のデータ、および、パリティをもとに、故障状態となったディスク装置と同一のデータ、および、パリティをスペアディスク装置に復元する。復元後は、スペアディスク装置が、故障状態となったディスク装置にかわりに動作する。

【0 0 0 4】

さらに、ディスク装置が故障状態となってからデータ、および、パリティの復元を行うと、レイド（R A I D）グループを構成する全ディスク装置にアクセスが発生し、オンラインの性能が低下することから、あらかじめ故障状態となりそうなディスク装置を予測しておき、故障状態となってアクセスができなくなる前にデータを予め各々対となっているスペアディスク装置にコピーし、スペアディスク装置によって運用を続ける技術がある。ディスク装置のエラー発生回数が規定値を超えた場合に、データをスペアディスク装置にコピーし、スペアディスクにデータを復元する技術が、特許文献 1 に開示されている。

【0 0 0 5】

さらに、従来のアレイ型ディスク装置では、予防保守等の理由によりディスク装置のスペアディスク装置へのデータ移行に際し、移行元のディスク装置で、データ読取り（リード）障害が多発した場合、移行元のディスク装置からデータリードを試み、データリード障害を検出後、アレイ型ディスク装置のデータ回復機能を用いて冗長性を持つディスク装置より移行元のデータを復元するというフローとなるため、ホストコンピュータからのデータリード要求のレスポンスの低下が予想される。このレスポンス低下を回避するため、移行元のディスク装置においてデータリードエラーが多発した場合、移行元のディスク装置をアレイ型ディスク装置から切り離し、アレイ型ディスク装置のデータ回復機能を用いて、冗長性を持つディスク装置より、移行元のデータを復旧する方式のみを用いてホストコンピュータのデータ読み出し要求に対応するという処理が一般的であった。

【0 0 0 6】

ところが、年々ディスク装置の容量は増加し、冗長性を持つアレイ型ディスク装置にもデータリード障害が発生する確率もそれに比例して増加するという問題が生じている。ま

た、冗長性を持つアレイ型ディスク装置にデータリード不能な部分があった場合、移行元のデータを復元することができず、結果として、データを失ってしまうという問題も発生する。

【0 0 0 7】

【特許文献 1】特開平 8 - 1 4 7 1 1 2 号公報

【0 0 0 8】

【非特許文献 1】"A Case for Redundant Arrays of Inexpensive Disks (RAID)", David A. Patterson, Garth Gibson, and Randy H. Katz, Computer Science Division Department of Electrical Engineering and Computer Sciences, University of California Berkeley

【発明の開示】

【発明が解決しようとする課題】

【0 0 0 9】

冗長なディスク装置即ちディスクアレイグループを 1 台分備えたアレイ型ディスク装置構成の場合、ディスク装置 1 台の故障の際にアレイ型ディスク装置の冗長性を利用してデータの回復することができるが、1 台のディスク装置が故障している状態でもう一台のディスク装置からの読み出しができなくなるとディスク 2 重障害となり、データを消失する。

【0 0 1 0】

アレイ型ディスク装置のデータ回復処理はオンライン処理と併行して行うことが一般的であり、年々ディスク装置の容量自体も増加しているため、データ回復処理時間が伸び、回復中にもう 1 台のディスク装置が故障する確率が上がる傾向にある。また、ディスク装置の容量の増加に伴い、データ回復時のディスク装置からの読み出し時間も増加し、回復不能なビットエラーが発生する確率も上がってきている。以上のことから、ディスク装置 2 重障害となる確率が増加する傾向にある。

【0 0 1 1】

ディスク装置へのアクセスができなくなる前に、データをスペアディスク装置にコピーしておく従来技術の場合、スペアディスク装置にコピーを開始する契機としてのエラー発生回数規定値を高くしておく、潜在的な故障の可能性を低く見えてしまうことになり、2 重障害となる確率が高くなる。また、エラー発生回数規定値を低くしておく、スペアディスク装置の使用が頻度が高くなり、スペアディスク装置のコストが高くなる。

【0 0 1 2】

また、ディスク装置が故障状態になったとアレイ型ディスク装置が判定した場合に、アレイ型ディスク装置のディスクアレイグループを形成する他のディスク装置のデータ、および、パリティをもとに、故障状態となったディスク装置と同一のデータ、および、パリティをスペアディスク装置に復元しようとアレイ型ディスク装置が試みるが、データ復元中に別のディスク装置において読み出しができないデータがあると、そのデータに関わるパリティグループのデータが復元できなくなってしまう、2 重障害となってしまうという問題がある。

【0 0 1 3】

また、アレイ型ディスク装置のディスクアレイグループを構成するディスク装置のうち、エラー発生回数が規定値に到達しているものはないながらも、複数のディスク装置のエラー発生回数が規定値に近くなっていて、潜在的にアレイ型ディスク装置のディスクアレイグループを構成するディスク装置のうち複数台が同時に故障するディスク 2 重障害となる可能性が高い場合がある。エラー発生回数をもとにスペアディスク装置へのコピーを開始する従来の技術では、上記の潜在的な 2 重障害の回避に対応できないという問題がある。

【0 0 1 4】

以上のように、従来の技術ではアレイ型ディスク装置を構成するディスク装置のうち複数台が同時に故障する 2 重障害への対応ができないケースがある。

【0015】

本発明の第一の目的は、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、スペアディスク装置のコストを上げることなく、ディスク2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することである。

【0016】

本発明の第二の目的は、アレイディスクを形成する1台のディスク装置が故障状態になったアレイ型ディスク装置において、ディスク装置2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することである。

【0017】

本発明の第三の目的は、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、アレイ型ディスク装置を形成する複数のディスク装置の故障ポテンシャルが高くなっている状態において、ディスク2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することである。

【0018】

本発明の第四の目的は、冗長性のあるアレイ型ディスク装置構成におけるディスク装置のスペアディスク装置へのデータ移行に際し、ホストコンピュータへのI/Oレスポンスを低下されることなく、かつデータを失うことなくデータ移行を完了されるアレイ型ディスク装置を提供することにある。

【0019】

さらには、上記四つの目的を達成するアレイ型ディスク装置を駆動する制御プログラム、制御方法、およびデータ移行方法を提供することにある。

【課題を解決するための手段】**【0020】**

本発明においては、上記目的を達成するために、複数のディスク装置を有するアレイ型ディスク装置において、少なくとも1台はスペアディスク装置とし、前記アレイ型ディスク装置は、前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値レベル1を超えた場合に前記ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、前記ディスク装置のエラー発生回数が前記規定値レベル1よりも大きい規定値レベル2を超えた場合に前記ディスク装置の閉塞開始を指示し、該ディスク装置で行っていた処理の前記スペアディスク装置への移行を指示するエラー監視部と、前記ディスク装置と前記スペアディスク装置とのミラーリングを行うミラー部と、前記ディスク装置の閉塞と前記移行とを行う閉塞移行部とを備えるようにした。

【0021】

また、前記アレイ型ディスク装置は、前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に、該ディスク装置と前記スペアディスク装置とのミラーリング開始を指示し、ミラーリングをしていないディスク装置のエラー発生回数がミラーリングをしているディスク装置のエラー発生回数を超えた場合、該スペアディスク装置のミラーリングを解除し、ミラーリングをしていないディスク装置とミラーリングを解除したスペアディスク装置とのミラーリングを開始するよう指示し、前記ディスク装置と前記スペアディスク装置とのミラーリングを行うようにした。

【0022】

さらに、前記アレイ型ディスク装置は、前記ディスク装置のエラー発生状況を監視し、前記ディスク装置のエラー発生回数が規定値を超えた場合に前記ディスク装置のステータスが仮閉塞状態となるよう指示するエラー監視部と、ディスクアレイグループを構成するディスク装置が仮閉塞状態となった場合、仮閉塞となったディスク装置のデータをディスクアレイグループを構成する他のディスク装置からスペアディスク装置に復旧するデータ復旧部とを備え、前記データ復旧部は、データ復旧中にディスクアレイグループを構成する他のディスク装置からの読み出しができない場合、仮閉塞状態のディスク装置からの読み出しを行うことでデータ復旧を行うようにした。

【0023】

さらにまた、複数のディスク装置を有するアレイ型ディスク装置のディスク装置間のデータ移行に際し、移行元ディスク装置からのデータ読み込みエラー発生回数を記憶し、エラー発生回数が規定値に達するまでは移行元のデータをディスク装置からデータを読み込み、エラー発生回数が規定値に達した場合、ディスクアレイグループを構成するデータディスク装置からのデータ読み込みに切り替え、この時ディスクアレイグループを構成するデータディスク装置からのデータ読み込みがエラーになり、データ復旧できない場合に、移行元のディスク装置からのデータ読み込みを行うようにした。

【0024】

また、前記アレイ型ディスクアレイ型ディスク装置は、前記ディスク装置が形成するディスクアレイグループを1単位としてディスク装置のエラー発生状況を監視し、エラー発生回数が規定値を超えた場合、該ディスク装置のデータをスペアディスク装置に移行することを指示し、前記ディスクアレイグループの複数の前記ディスク装置のエラー発生回数が、規定値よりも小さく設定した補助規定値に達した場合、上記規定値をより小さな値に動的に変更し、前記移行指示を受けてデータ移行を行うようにした。

【発明の効果】

【0025】

本発明によれば、ディスクアレイ (RAID) グループを構成するディスク装置のうち複数台が同時に故障する2重障害の発生を抑えることができる。

【0026】

スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、あらかじめスペアディスク装置にミラーリングしておき、スペアディスク装置をミラーリングしていなかったディスク装置へのスペアとして利用できるのもので、スペアディスク装置のコストを上げることなく、ディスク2重障害の確率を下げるることができるという効果がある。

【0027】

また、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、エラー発生回数が少ないうちから、エラー発生回数の多いディスク装置に対してミラーリングを行っておき、ミラーリングを組むディスク装置をエラー発生回数に応じて、ダイナミックに切り替えていくことで、第2段階の規定値に達したときに即時にスペアディスク装置への切り替えができるという効果がある。

【0028】

また、ディスクアレイ (RAID) グループを形成する1台のディスク装置が故障状態になったディスクアレイシステムにおいて、ディスク2重障害の確率を下げるることができるという効果がある。

【0029】

また、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、ディスクアレイ (RAID) グループを形成する複数のディスク装置の故障ポテンシャルが高くなっている状態において、ディスク装置2重障害の確率を下げるることができるという効果がある。

【0030】

さらにまた、大容量アレイ型ディスク装置におけるディスク装置間のデータ移行の際、移行元のディスク装置を完全に切り離すことなく使い続けることにより、冗長データによるデータの復旧方式と、移行元のディスク装置からのリード方式のハイブリッド方式により、データを失うことなく移行先のディスク装置にデータ移行が可能となる効果を奏する。

【発明を実施するための最良の形態】

【0031】

(第1の実施形態)

本発明の第1の実施形態は、発明が解決しようとする課題の第1の目的を達成するためのものである。

すなわち、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、ス

ペアディスク装置のコストを上げることなく、ディスクドライブあるいは単にドライブとも称されるディスク装置の2重障害の確率を低減する信頼性の高いアレイ型ディスク装置を提供することである。

(1) 構成の説明

本発明の第1の実施形態のシステム構成を図1～3を用いて説明する。図1において、100はホストコンピュータ、123はアレイ型ディスク装置、200はアレイ型ディスク装置の管理制御部、310はディスク装置群、500は管理端末である。

【0032】

アレイ型ディスク装置123、ホストコンピュータ100、管理制御部200、ディスク装置群310、および、管理端末500はそれぞれ図のように結線する。

【0033】

アレイ型ディスク装置123は管理制御部200として以下を含む。201は管理制御部200の制御を実施するCPU、202はメモリ、203はユーザのデータをバッファリングするキャッシュ、204はホストとのデータの送受信を行うホストI/F、205はディスク装置群310との送受信を行うドライブI/F、207は管理端末500との制御情報の送受信を行う管理I/Fであり、それぞれを結線する。メモリ202は、ディスクアレイを制御するレイド(RAID)制御部210と、ディスク装置群310の管理を行うディスク装置管理部230と、ディスク装置群310の運用パラメータや運用状況などディスク装置情報を記録するディスク装置管理テーブル240と、管理端末500からの入力を受けてディスク装置情報を設定するディスク装置情報設定部250と、管理端末500への出力としてディスク装置情報を通知するディスク装置情報通知部260とを有する。

【0034】

ディスク装置群310は、ディスク装置301～307とからなる。ディスク装置301～305は、従来の技術で述べたようなディスクの並列動作と冗長性の付加により、性能と信頼性を強化したディスクアレイグループを構成しており、この状態をディスク装置301～305の組でレイド(RAID)グループとなるディスクアレイグループを構成していると呼ぶこととする。ディスク装置306、307は、ディスクアレイ(RAID)グループを構成するディスクが故障したときに替わりにディスクアレイグループに組み入れるためのスペアディスク装置である。

【0035】

管理端末500は、ユーザからのディスク装置301～305への設定を入力する入力部510と、ディスク装置301～305の情報をユーザに示す出力部520とからなる。入力部510からは、ディスク装置管理テーブル240へのディスク装置運用パラメータを入力する。出力部520は、ディスク装置管理テーブル240のディスク装置運用状況を出力表示する。

【0036】

図2は、ディスク装置管理テーブル240である。パラメータとしては、ディスク装置の識別番号を表す「ディスク装置No.」、ディスク装置の累積エラー回数を格納する「エラー回数カウンタ」、ディスク装置の累積エラー回数の指標として第1段階の値を示す「エラー回数規定値レベル1」、ディスク装置の累積エラー回数の指標として第2段階の値を示す「エラー回数規定値レベル2」、スペアディスク装置として運用されていることを示す「スペアビット」、ディスク装置の運用状況を示す「ディスク装置状況」、ディスク装置障害に対応するために使うスペアディスク装置との関連付けを示す「ペアディスク装置」を有する。

【0037】

「エラー回数規定値レベル1」には、対象となるディスク装置のエラー回数が累積し障害が発生する可能性が高くなった場合に、スペアディスク装置とのミラーリングを開始する契機を示す値を設定する。「エラー回数規定値レベル2」には、「エラー回数規定値レベル1」の値よりも高い値を設定し、対象となるディスク装置のエラー回数が累積し運用

が継続できないとみなせる状況になったと判定する値とし、ディスク装置を閉塞させ、スペアディスク装置とのミラーリングを終了する契機を示す値を設定する。「スペアビット」には、該当ディスク装置がスペアディスク装置である場合はYESを、そうでない場合はNOを設定する。「エラー回数規定値レベル1」、「エラー回数規定値レベル2」、「スペアビット」の設定は、ユーザが管理端末500の入力手段510を用いて設定する。

【0038】

「ディスク装置状況」には、ディスク装置の運用状況として異常がないことを示す「正常」、エラー回数カウンタの値が「エラー回数規定値レベル1」に達し、スペアディスク装置とのミラーリングを実施していることを示す「ミラー」、エラー回数カウンタの値が「エラー回数規定値レベル2」に達し、該当ディスク装置を運用継続不可とみなしたことを示す「閉塞」のパラメータを設定する。「ペアディスク装置」には、ミラーリングするペアとなるディスク装置の「ディスク装置No.」を設定する。ディスク装置管理テーブル240の各パラメータ値は、ユーザ指示により管理端末500の出力手段520に出力表示する。

【0039】

図3は、ディスク装置管理部230である。231は、ディスク装置のエラー発生状況を監視し、ディスク装置のエラー発生回数が「エラー回数規定値レベル1」を超えた場合にディスク装置とスペアディスク装置のミラーリング開始を指示し、「エラー回数規定値レベル2」を超えた場合はミラーリングを終了するように指示するエラー監視部である。232は、ディスク装置のエラー発生回数をカウントし、ディスク装置管理テーブル240の「エラー回数カウンタ」に積算したエラー発生回数を設定するエラーカウント部である。233は、管理端末500を使ってユーザが指定したパラメータをディスク装置管理テーブル240に設定するエラー回数規定値設定部である。234は、エラー監視部231の指示により、ディスク装置の運用状況をディスク装置管理テーブル240に設定するディスク状況設定部である。235は、あるディスク装置とスペアディスク装置へのアクセスをミラーリングするミラー部である。236はディスク装置の閉塞と該装置が行っていた処理のスペアディスク装置への移行を指示する閉塞移行監視部である。237は閉塞監視部の指示によりディスク装置の閉塞・移行を行う閉塞移行部である。

【0040】

以上が、本実施形態のアレイ型ディスク装置のシステム構成である。

(2) 予防スペアコピー動作

従来の技術では、ディスク装置のエラー発生回数を監視し、ある規定値に達した場合にスペアディスク装置にコピーし、該当ディスク装置を閉塞させるのに対し、本実施形態では、規定値を2段階設け、第1段階の規定値レベル1に達した場合にスペアディスク装置とのミラーリングを開始する。その際、該当ディスク装置は閉塞させずに運用を継続する。第2段階の規定値レベル2に達した場合にミラーリングを解除し、該当ディスク装置を閉塞させスペアディスク装置にて運用を継続する。

【0041】

次に、予防スペアコピー動作を図4のフローチャートを用いて説明する。

【0042】

前提として、各ディスク装置301～307のエラー発生状況は、エラーカウント部232によってカウントし、ディスク装置管理テーブル240に継続的に設定されているものとする。図4は、ディスクアレイグループ(RAID)グループを構成するディスク装置301～305に対して、独立に実施するものとする。

【0043】

まず、エラー監視部231が、監視の対象としている該当ディスク装置のディスク装置管理テーブル240の「エラー回数カウンタ」の値が「エラー回数規定値レベル1」に達したかどうかを判定する(ステップ1001)。達していない場合は、引き続きステップ1001を繰り返す。達していた場合は、「スペアビット」がYesとなっているディスク装置を探索し、スペアディスク装置を選択する(ステップ1002)。その後、エラー

監視部 231 は、選択したスペアディスク装置の「ペアディスク装置」に該当ディスク装置のディスク装置番号を設定し（ステップ 1003）、対象となるディスク装置の「ペアディスク装置」にスペアディスク装置の番号を設定する（ステップ 1004）。次に、エラー監視部 231 は、対象となるディスク装置とスペアディスク装置の「ディスク装置状況」をミラーのステータスに設定し（ステップ 1005）、ミラー部 235 に指示を送り、該当ディスク装置とスペアディスク装置のミラーリングを開始する（ステップ 1006）。

【0044】

ここで、図 2 にディスク装置管理テーブル 240 の設定状況の例を示す。「ディスク装置 No.」として 0～4 にてディスクアレイ（RAID）グループを構成するディスクアレイにおいて、「ディスク装置 No.」の 4 のディスク装置が「エラー回数規定値レベル 1」である「50」という値を超えた「エラー回数カウンタ」の値 60 となっている。この状態はすでにスペアディスク装置である「ディスク装置 No.」の 5 とのミラーリングが開始された状況であり、「ディスク装置 No.」の 4 のディスク装置の「ディスク装置状況」はミラー、「ペアディスク装置」は No. 5 となっており、「ディスク装置 No.」の 5 のディスク装置の「ディスク装置状況」はミラー、「ペアディスク装置」は No. 4 となっている。

【0045】

図 4 に戻り、次のステップとして、エラー監視部 231 が、監視の対象としている該当ディスク装置のディスク装置管理テーブル 240 の「エラー回数カウンタ」の値が「エラー回数規定値レベル 2」に達したかどうかを判定する（ステップ 1007）。達していない場合は、引き続きステップ 1007 を繰り返す。達していた場合は、閉塞移行監視部が閉塞開始およびスペアディスク装置への移行開始を指示し、対象となるディスク装置の「ディスク装置状況」を閉塞のステータスに、スペアディスク装置の「ディスク装置状況」を正常のステータスに設定し（ステップ 1008）、ミラー部 235 に指示を送り、該当ディスク装置とスペアディスク装置のミラーリングを終了し、対象ディスク装置に行っていた処理をスペアディスク装置に移行する（ステップ 1009）。上記の閉塞及び移行は閉塞移行部により行う。スペアディスク装置がどのディスクから移行したかは「ペアディスク装置」の値を参照すればよい。

【0046】

以上が、予防スペアコピー動作である。

(3) 効果

従来の技術では、ディスク装置のエラー発生回数を監視し、ある規定値に達した場合にスペアディスク装置にコピーし、該当ディスク装置を閉塞させるのに対し、本実施形態では、規定値を 2 段階設け、第 1 段階の規定値に達した場合にスペアディスク装置とのミラーリングを開始する。その際、該当ディスク装置は閉塞させずに運用を継続する。第 2 段階の規定値に達した場合にミラーリングを解除し、該当ディスク装置を閉塞させスペアディスク装置にて運用を継続する。

【0047】

また、ミラーリングしているだけであるため、仮に該当ディスク装置以外のディスク装置で、第 2 段階の規定値を超えるようなエラー発生状況となった場合、該当ディスク装置のミラーリングを解除し、スペアディスク装置を他のディスク装置のスペアとして使用することも可能である。

【0048】

例えば、図 2 のディスク装置管理テーブル 240 の例が示すように、「ディスク装置 No.」の 4 のディスク装置が「エラー回数規定値レベル 1」である「50」という値を超えた「エラー回数カウンタ」の値 60 となっている場いいで、「ディスク装置 No.」4 のディスクと、「ディスク装置 No.」の 5 のディスク装置が 60 となっている場合で、「ディスク装置 No.」の 5 のディスク装置がミラーリングされている状態を想定する。この状態で、「ディスク装置 No.」の 0 のディスク装置の「エラー回数カウンタ」の値が、「エラ

一回数規定値レベル2」である「90」という値を超えた場合、エラー監視部231が「ディスク装置No.」の4と5のミラーリングを解除し、「ディスク装置No.」の5のディスク装置を「ディスク装置No.」の0のディスクのスペアとして使用することができる。「ディスクNo.」の0のディスクはエラー発生頻度が高くなり、故障状態となりそうなディスク装置であり、故障状態となる前にデータをスペアディスク装置にコピーする。

【0049】

このように、他のディスク装置のスペアディスク装置としての転用が可能であるため、第1段階の規定値は、従来の技術で規定していた値よりも、低い値とすることが可能で、ディスクの2重障害に対する耐性を向上できる。また、スペアディスクの転用が可能であるため、第1段階で閉塞させる従来の技術と比較して、スペアディスク装置のコストを抑止できる。

【0050】

また、第1段階でミラーリングしておくので、第2段階の規定値に達したときに即時にスペアディスク装置への切り替えが可能であり、ディスク装置の2重障害に対する耐性を向上できる。

【0051】

以上、本実施形態により、スペアディスク装置に予防コピーしておくディスクアレイにおいて、スペアディスク装置のコストを上げることなく、ディスク2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することができる。

【0052】

(第2の実施形態)

第2の実施形態は、第1の実施形態と同様に、発明が解決しようとする課題の第1の目的を解決するためのものである。すなわち、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、スペアディスク装置のコストを上げることなく、ディスク装置の2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することである。

(1) 構成の説明

本発明の第2の実施形態のシステム構成を説明する。簡単化のため、第1の実施形態との違いについてのみ述べる。システム構成は第1の実施形態と同様に図1である。

ディスク装置群310は、ディスク装置301から307とからなる。ディスク301～305は、従来の技術で述べたようなディスク装置の並列動作と冗長性の付加により、性能と信頼性を強化したディスクアレイを構成しており、この状態をディスク装置301から305の組でディスクアレイ(RAID)グループを構成していると呼ぶこととする。ディスク装置306、307はディスクアレイ(RAID)グループを構成するディスク装置が故障したときに替わりにディスクアレイ(RAID)グループに組み入れるためのスペアディスク装置であるが、第1の実施形態との違いはエラー発生回数が少ないことから、エラー発生回数の多いディスク装置に対してミラーリングを行っておくことである。ミラーリングを行うスペアディスク装置は全スペアディスク装置、もしくは、2台以上のスペアディスク装置があることが望ましいが1台のスペアディスクでもよい。また、ミラーリングしているディスク装置以外のディスク装置で、ミラーリングをしているディスク装置のエラー発生回数を超える状況になった場合には、ミラーリングをしているディスク装置のうちエラー発生回数が最も低いディスク装置のミラーリングを解除し、解除したスペアディスク装置をエラー発生回数が高くなったディスク装置に対するミラーリングを行うスペアディスク装置として使用する。ミラーリングを組むディスク装置をダイナミックに切り替えていくことから、この動作をダイナミックミラーリング動作と呼ぶこととする。

【0053】

図5は、第2の実施形態のディスク装置管理テーブル240であり、パラメータは、第1の実施形態の図2と同様である。第1の実施形態との違いは、「エラー回数規定値レベル1」には、対象となるディスク装置のエラー回数が累積し障害が発生する確率が高くな

った場合に、全ディスク装置の「エラー回数カウンタ」を調査し、スペアディスク装置を「エラー回数カウンタ」の値の高い方のディスク装置とミラーリングを開始する契機を示す値を設定する。

【0054】

「ディスク装置状況」には、ディスク装置の運用状況として異常がないことを示す「正常」、スペアディスク装置とのミラーリングを実施していることを示す「ミラー」、エラー回数カウンタの値が「エラー回数規定値レベル2」に達し、該当ディスク装置を運用継続不可とみなしたことを示す「閉塞」のパラメータを設定する。

【0055】

第2の実施形態においては、ディスク装置管理部230は図3であり、231は、ディスク装置のエラー発生状況を監視し、ディスク装置のエラー発生回数が「エラー回数規定値レベル1」を超えた場合、全ディスク装置の「エラー回数カウンタ」を調査し、スペアディスク装置を「エラー回数カウンタ」の値の高い方のディスク装置とミラーリングを開始し、「エラー回数規定値レベル2」を超えた場合はミラーリングを終了するよう指示するエラー監視部である。

以上が、本実施形態のシステム構成である。

【0056】

(2) ダイナミックミラーリング動作

従来の技術では、ディスク装置のエラー発生回数を監視し、ある規定値に達した場合にスペアディスク装置にコピー(ミラーリング)し、該当ディスク装置を閉塞させるのに対し、本実施形態では、エラー発生回数が少ないうちから、エラー発生回数の多いディスク装置に対しミラーリングを行っておき、ミラーリングを組むディスク装置をエラー発生回数に応じて、ダイナミックに切り替えていく。

【0057】

次に、ダイナミックミラーリング動作を図6-1及び図6-2のフローチャートを用いて説明する。前提として、各ディスク装置301から307のエラー発生状況は、エラーカウンタ部232によってカウントし、ディスク装置管理テーブル240に継続的に設定されているものとする。

【0058】

まず、エラー監視部231が、ディスク装置管理テーブル240の「エラー回数カウンタ」の値が「エラー回数規定値レベル1」に達したディスク装置があるかどうかを判断する(ステップ1501)。この場合、「エラー回数規定値レベル1」に達したディスク装置はどのディスク装置であってもよい。「エラー回数カウンタ」の値が「エラー回数規定値レベル1」に達したディスク装置がない場合は、引き続きステップ1501を繰り返す。

【0059】

「エラー回数規定値レベル1」に達したディスク装置がある場合は、全ディスク装置の「エラー回数カウンタ」の値を調査する(ステップ1502)。次に、エラー監視部231は、「スペアビット」がYESとなっているディスク装置を探索し、「ミラー状況」がミラーとなっていないディスク装置、即ちペアを組んでいないスペアディスク装置があるかどうかを判定する(ステップ1503)。

【0060】

ペアを組んでいないスペアディスク装置がある場合は、エラー監視部231が、ペアを組んでいないディスク装置のうち、「エラー回数カウンタ」の値が最大のものをペアリング対象に選定し(ステップ1504)、スペアディスク装置の「ペアディスク装置」に対象となるディスク装置番号を設定し(ステップ1505)、対象となるディスク装置の「ペアディスク装置」にスペアディスク装置番号を設定し(ステップ1506)、対象となるディスク装置とスペアディスク装置の「ディスク装置状況」をミラーのステータスに設定し(ステップ1507)、ミラー部235にミラー開始を指示し(ステップ1508)、ステップ1503に戻る。

ペアを組んでいないスペアディスク装置がない場合は、ステップ1509に移行する。

【0061】

ここで、図5にディスク装置管理テーブル240の設定状況の例を示す。「ディスク装置No.」として0～4にてレイド(RAID)グループを構成するディスクアレイにおいて、「ディスク装置No.」の2のディスクが「エラー回数規定値レベル1」である「30」という値を超えた「エラー回数カウンタ」の値35となっている。この状態はすでにステップ1509まで移行した状況であり、スペアディスク装置である「ディスク装置No.」の5と「ディスク装置No.」の2のミラーリングが開始された状況であり、「ディスク装置No.」の2の「ディスク装置状況」はミラー、「ペアディスク装置」はNo.5となっており、「ディスク装置No.」の5のディスク装置の「ディスク装置状況」はミラー、「ペアディスク装置」はNo.2となっている。また、2番目に「エラー回数カウンタ」の値が大きかった「ディスク装置No.」の4のディスク装置とペアディスク装置である「ディスク装置No.」の6のミラーリングが開始された状況であり、「ディスク装置No.」の4のディスク装置の「ディスク装置状況」はミラー、「ペアディスク装置」はNo.6となっており、「ディスク装置No.」の6のディスクの「ディスク装置状況」はミラー、「ペアディスク装置」はNo.4となっている。

【0062】

図6-2に戻り、次のステップ1509として、エラー監視部231が、ペアを組んでいないディスク装置のうち、ペアを組んでいるディスク装置を超える「エラー回数カウンタ」の値となったディスク装置があるかどうか判定する(ステップ1509)。

【0063】

該当のディスク装置がある場合は、エラー監視部231が、ペアを組んでいないディスク装置のうち、ペアを組んでいるディスク装置を超える「エラー回数カウンタ」の値となったディスクをペアリング対象に選定し(ステップ1510)、ペアを組んでいるディスクのうち、「エラー回数カウンタ」の値が最小のディスクにおけるペアリングを解除し(ステップ1511)、ペアリングを解除したスペアディスク装置の「ペアディスク装置」に対象となるディスク装置番号を設定し(ステップ1512)、対象となるディスク装置の「ペアディスク装置」にスペアディスク装置番号を設定し(ステップ1513)、対象となるディスク装置とスペアディスク装置の「ディスク装置状況」をミラーのステータスに設定し(ステップ1513)、ミラー部235にミラー開始を指示し(ステップ1515)、ステップ1509に戻る。

【0064】

ここで、図5のディスク装置管理テーブル240の設定状況の例を用いて、ステップ1509からステップ1515を説明する。スペアディスク装置である「ディスク装置No.」の5と「ディスク装置No.」の2のミラーリングが実施され、スペアディスク装置である「ディスク装置No.」の6と「ディスク装置No.」の4のミラーリングが実施されている状況を示している。

【0065】

この状況において、例えば、「ディスク装置No.」の0のディスク装置の「エラー回数カウンタ」の値が、ミラーリングしている「ディスク装置No.」のディスク装置の値を超える25となった場合を想定する。この場合、ステップ1509でいうところのYesの状況であり、次のミラーリング対象を「ディスク装置No.」の0のディスク装置とし、ミラーリングしているディスク装置のうち「エラー回数カウンタ」の値が最小の「ディスク装置No.」の4のペアリングを解除し、ペアを解除したスペアディスク装置である「ディスク装置No.」の6と、新規にミラーリングの対象となった「ディスク装置No.」の0のディスク装置のミラーリングを実施するように動作する。

【0066】

図6-2に戻り、ステップ1509においては、ペアを組んでいないディスク装置のうち、ペアを組んでいるディスク装置を超える「エラー回数カウンタ」の値となったディスク装置がない場合はエラー監視部231が、監視の対象としているディスク装置の「エラー回数カウンタの値」が「エラー回数規定値レベル2」に達したかどうか判定する(ステ

ップ1516)。達していない場合は、ステップ1509に戻る。達していた場合は、対象となるディスク装置の「ディスク装置状況」を関連のステータスに、スペアディスク装置の「ディスク装置状況」を正常のステータスに設定し(ステップ1517)、ミラー部235に指示を送り、該当ディスク装置とスペアディスク装置のミラーリングを終了し、対象ディスク装置に行っていた処理をスペアディスク装置に移行し(ステップ1518)、ステップ1509に戻る。スペアディスク装置がどのディスク装置に移行したかは「ペアディスク装置」の値を参照すればよい。

以上がダイナミックミラーリングの動作である。

【0067】

なお、「エラー回数規定値レベル1」の値を0にしておき、当初からステップ1502移行のダイナミックミラーリング動作を実施してもよい。

また、ステップ1509における判定基準を、ペアを組んでいないディスク装置のうち、ペアを組んでいるディスク装置中の「エラー回数カウンタ」の最大値を超えたディスク装置があるかどうかの判定としてもよい。または、ペアを組んでいるディスク装置の「エラー回数カウンタ」の値から導かれる中間値、平均値等の値を超えたディスク装置があるかどうかの判定としてもよい。

【0068】

(3) 効果

従来の技術では、ディスク装置のエラー発生回数を監視し、ある規定値に達した場合にスペアディスク装置にコピーし、該当ディスク装置を閉塞させるのに対し、本実施形態では、エラー発生回数が少ないうちから、エラー発生回数の多いディスク装置に対してミラーリングを行っておき、ミラーリングを組むディスク装置をエラー発生回数に応じて、ダイナミックにきりかえていくため、第2段階の規定値に達したときに即時にスペアディスク装置への切り替えができる確率が上がり、ディスク装置の2重障害に対する耐性を向上できる。

【0069】

なお、ダイナミックミラーリングを行うディスク装置を1つのアレイディスク (RAID) グループに対して行うよう記載したが、アレイ型ディスク装置全体にあるアレイディスク (RAID) グループに対して、アレイ型ディスク装置内にある全スペアディスク装置を用いてダイナミックミラーリングを実施してもよい。

【0070】

(第3の実施形態)

第3の実施形態は、発明が解決しようとする課題の第2の目的を解決するためのものである。

すなわち、ディスクアレイ (RAID) グループを形成する1台のディスク装置が故障状態になったアレイ型ディスク装置において、ディスク2重障害の確率を低減する信頼性の高いアレイ型ディスク装置を提供することである。

【0071】

(1) 構成の説明

本発明の第3の実施形態の装置構成を図7～9を用いて説明する。簡単化のため、第1の実施形態との違いについてのみ述べる。図7においては図1の構成に加え、ディスク装置が閉塞し、ディスクアレイ (RAID) グループを構成する他のディスク装置からスペアディスク装置にデータを復元するデータ復旧部270をメモリ202に設置する。

【0072】

また、図7のディスク装置管理テーブル240が所持するパラメータは、図2の場合と比較してエラー回数規定値レベル2を除いたものである。また、以下の点で格納内容が異なる。

【0073】

「エラー回数規定値レベル1」には、対象となるディスク装置のエラー回数が累積し障害が発生する可能性が高くなった場合に、スペアディスク装置にコピーを開始する契機を

示す値を設定する。コピー終了後、対象となるディスク装置の処理はスペアディスク装置に移行するが、データ復旧部 270 が実施する対象ディスク装置からの読み出しは許可する。

【0074】

「ディスク装置状況」には、ディスク装置の運用状況として異常がないことを示す「正常」、エラー回数カウンタの値が「エラー回数規定値レベル1」に達し、スペアディスク装置にコピーしている状態である「コピー」、スペアディスク装置にコピーが終了し、データ復旧手段 270 が実施する対象ディスク装置からの読み出しは許可する状態である「仮閉塞」、コピーが終了した後の「閉塞」、アレイディスク (RAID) グループを構成する他のディスク装置からスペアディスク装置にデータを復元する処理」には、コピーするペアとなるディスク装置の「ディスク装置 No.」を設定する。

【0075】

また、図 9 は、第 3 の実施形態のディスク装置管理部 230 であり、図 3 のミラー部 235 に替わり、236 のコピー部を有する。231 のエラー監視部は、ディスクのエラー発生状況を監視し、ディスク装置のエラー発生回数が「エラー回数規定値レベル1」を超えた場合にディスク装置からスペアディスク装置へのコピー開始を指示し、コピー中は仮閉塞のステータスにし、コピー終了後は、閉塞のステータスにする。236 のコピー部は、あるディスクのデータをスペアディスクにコピーする。

以上が、本実施形態のシステム構成である。

【0076】

(2) セクタ障害復旧動作

本実施形態では、あるセクタの読み出しができなくなり、ディスクアレイ (RAID) グループを構成する他のディスク装置からスペアディスク装置にデータを復旧することになった状況において、さらに、ディスクアレイ (RAID) グループを構成する他のディスク装置のあるセクタが読み出せなくなるディスク 2 重障害のケースでのデータ復旧の可能性を向上させる。あるセクタの読み出しができなくなったディスク装置はデータ復旧部 270 が実施する読み出しについては許可する仮閉塞状態としておく。

【0077】

次に、セクタ障害復旧動作を図 10 のフローチャートを用いて説明する。

前提として、各ディスク装置 301～307 のエラー発生状況は、エラーカウント手段 232 によってカウントし、ディスク装置管理テーブル 240 に継続的に設定されているものとする。図 10 は、ディスクアレイ (RAID) グループを構成するディスク装置 301～305 に対して、独立に実施するものとする。また、ディスクアレイ (RAID) グループを構成する「ディスク装置 No.」の 4 は、エラー回数が多くなっていると同時にあるセクタが読み出せなくなっており、エラー回数カウンタにかかわらず、仮閉塞状態としている。「ディスク装置 No.」の 0～3 を用いて、スペアディスク装置である「ディスク装置 No.」の 5 にディスクアレイ (RAID) の冗長性を使ってデータを復旧している状況を想定する。この状況において、「ディスク装置 No.」の 0 のあるセクタが読み出せなくなり、「ディスク装置 No.」の 4 の同一セクタからデータを読み出し、ディスクアレイ (RAID) グループの復旧をするものと想定する。

【0078】

まず、データ復旧部 270 が、「ディスク装置 No.」の 0～3 のデータをもとに、「ディスク装置 No.」の 5 のスペアディスク装置に、「ディスク装置 No.」の 4 相当のデータ復旧処理を開始する (ステップ 2001)。次に、データ復旧部 270 は、復旧が終了したかどうかを判定し (ステップ 2002)、終了した場合は復旧対象となる「ディスク装置 No.」の 4 の処理をスペアディスク装置に移行し (ステップ 2003)、処理を終了する (ステップ 2004)。終了していない場合は、復旧に使用している「ディスク装置 No.」の 0～3 のディスク装置にセクタが読み出せないセクタ障害があるかどうかを判定する (ステップ 2005)。セクタ障害がない場合は、引き続きステップ 2002 を繰り返す。セクタ障害がある場合は、仮閉塞となっている「ディスク装置 No.」の

4の同一セクタから読み出しを試みる(ステップ2006)。データ復旧部270は、読み出しが成功したかどうかの判定を行い(ステップ2007)、成功した場合は読み出したセクタの内容をもとに復旧処理を実施し(ステップ2008)、ステップ2002に戻る。失敗した場合は、対応セクタをデータロス扱いにし(ステップ2009)、ステップ2002に戻る。

【0079】

以上が、セクタ障害復旧動作である。

【0080】

(3) セクタ障害復旧動作時のライト動作

前提として、各ディスク装置301~307のエラー発生状況は、エラーカウント部232によってカウントし、ディスク装置管理テーブル240に継続的に設定されているものとする。図1は、ディスク装置301~305により構成するディスクアレイ(RAID)グループ全体に対して実施するものとする。また、ディスク装置301~305によりディスクアレイ(RAID)グループを構成しており、各ディスク装置内にはデータとパリティが格納されており、パリティとパリティを算出するためのデータの組をストライプセットと呼ぶこととする。

【0081】

図11において、まず、管理制御部200がホスト100からの書き込み要求を受け取ると、ディスクアレイ(RAID)制御部210は、書き込み先が仮閉塞しているディスク装置かどうかを判定する(ステップ2501)。

【0082】

書き込み先が仮閉塞ディスク装置の場合、ステップ2502以降の処理となる。ここで仮閉塞ディスク装置をディスク装置305、書き込むデータの同一ストライプセットのパリティが格納されているディスク装置をディスク装置301とする。まずディスクアレイ(RAID)制御部210は、仮閉塞しているディスク装置305とパリティが格納されているディスク装置301以外のディスク装置302から304から書き込むデータに対応する同一ストライプセットのデータを読み出す(ステップ2502)。次に、書き込みデータとステップ2502で読み出したデータの排他的論理和を算出し、新パリティを生成する(ステップ2503)。次に、書き込むデータを仮閉塞ディスク装置であるディスク装置305に書き込み(ステップ2504)、新パリティをパリティが格納されているディスク装置301に格納し(ステップ2505)、処理を終了する。

【0083】

書き込み先が仮閉塞しているディスク装置でない場合、ステップ2507以降の処理となる。ディスクアレイ(RAID)制御部210は、書き込むデータのストライプセットのパリティが仮閉塞しているディスク装置にあるかどうか判定する(ステップ2507)。

【0084】

パリティが仮閉塞しているディスク装置にある場合、ステップ2508以降の処理となる。ここで、仮閉塞しているディスク装置をディスク装置305、書き込むデータが格納されているディスク装置を301とする。まず、ディスクアレイ(RAID)制御部210は仮閉塞しているディスク装置305とデータが格納されているディスク装置301以外のディスク装置302空04から書き込むデータに対応する同一ストライプセットのデータを読み出す(ステップ2508)。次に、書き込みデータとステップ2508で読み出した同一ストライプセットのデータの排他的論理和を算出し、新パリティを生成する(ステップ2509)。次に、書き込むデータをディスク装置301に書き込み(ステップ2510)、新パリティをパリティが格納されている仮閉塞しているディスク装置であるディスク装置305に格納し(ステップ2511)、処理を終了する。

【0085】

パリティが仮閉塞しているディスク装置にない場合、ステップ2512以降の処理となる。ここで仮閉塞しているディスク装置をディスク装置305、書き込むデータが格納されているディスク装置をディスク装置301、同一ストライプセットのパリティが格納さ

れているディスク装置をディスク装置302とする。まず、ディスクアレイ (RAID) 制御部210は、書き込むデータの更新前のデータが格納されているディスク装置301から旧データを読み出し、更新前のパリティが格納されているディスク装置302から旧パリティを読み出す(ステップ2512)。次に、書き込みデータと、ステップ2512で読み出した旧データと旧パリティの排他的論理和を算出し、新パリティを生成する(ステップ2513)。次に、書き込むデータをディスク装置301に書き込み(ステップ2514)、新パリティをパリティが格納されているディスク装置302に格納し(ステップ2515)、処理を終了する。

【0086】

以上が、セクタ障害復旧中にホストより書き込み要求が来た場合のライト動作である。

なお、ディスクアレイ (RAID) の冗長性を利用してデータを復旧できることから、ステップ2504、ステップ2511での仮閉塞しているディスク装置への書き込みは省略してもよい。また、ステップ2504、ステップ2511での仮閉塞しているディスクへの書き込みに替えて、スペアコピーを実施しているスペアディスク装置に書き込むようにしてもよい。また、ステップ2504、ステップ2511での仮閉塞しているディスク装置への書き込みに加えて、スペアコピーを実施しているスペアディスク装置にも仮閉塞しているディスク装置に書き込み内容を書き込むようにしてもよい。

(4) 効果

本実施形態では、あるセクタの読み出しができなくなり、ディスクアレイ (RAID) グループを構成する他のディスク装置からスペアディスク装置にデータを復旧することになった状況において、さらに、ディスクアレイ (RAID) グループを構成する他のディスク装置のあるセクタが読み出せなくなるディスク装置2重障害のケースでのデータ復旧の可能性を向上させることができる。

【0087】

以上、本実施形態により、ディスクアレイ (RAID) グループを形成する1台のディスク装置が故障状態になったアレイ型ディスク装置において、ディスク装置2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することができる。

【0088】

なお、スペアディスク装置に予防コピーを行うことを前提として記載したが、予防コピーを行わないアレイ型ディスク装置においても、本実施形態は適用可能である。

【0089】

また、アレイ型ディスク装置の復旧開始の前提として、あるセクタ読み出しができなくなったディスクアレイ型ディスク装置の存在が復旧の契機であるとしたが、その他の条件であっても構わない。たとえば、エラー発生回数が規定値を超えたことにより、閉塞状態とみなしたことディスク装置の復旧を開始したという契機であってもよい。

【0090】

(第4の実施形態)

第4の実施形態は、発明が解決しようとする課題の第3の目的を解決するためのものである。

すなわち、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、ディスクアレイ (RAID) グループを形成する複数のディスク装置の故障ポテンシャルが高くなっている状態において、ディスク装置2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することである。

(1) 構成の説明

本発明の第4の実施形態のアレイ型ディスク装置構成を図12、13を用いて説明する。簡単化のため、第3の実施形態との違いについてのみ述べる。アレイ型ディスク装置構成図としては、第2の実施形態の図7と同一である。ただし、データ復旧手段270の機能として、データ復旧中にセクタ障害があった場合に、かわりとなるセクタを読み出す機能は持たなくてもよい。

【0091】

また、図12のディスク装置管理テーブル240は、所持するパラメータは図8と比較してエラー回数補助規定値を付加したものである。また、以下の点で図8と格納内容が異なる。

【0092】

「エラー回数規定値レベル1」には、対象となるディスク装置のエラー回数が累積し障害が発生する可能性が高くなった場合に、スペアディスク装置にコピーを開始する契機を示す値を設定する。コピー終了後、対象となるディスク装置の処理はスペアディスク装置に移行し、対象となるディスク装置は閉塞状態とする。「エラー回数補助規定値」には、「エラー回数規定値レベル1」の値よりも低い値を設定し、ディスクアレイ(RAID)グループを構成するディスク装置のうち、複数のディスク装置がこの値に達すると、潜在的に同時に障害を起こす危険性が高いことを意味する値とする。

【0093】

「ディスク装置状況」には、ディスク装置の運用状況として異常がないことを示す「正常」、エラー回数カウンタの値が「エラー回数規定値レベル1」に達し、スペアディスク装置にコピーしている状態である「コピー」、スペアディスク装置にコピーが終了した後の「閉塞」、ディスクアレイ(RAID)グループを構成する他のディスク装置からスペアディスク装置にデータを復元する処理を実行中であることを示す「復旧中」の各パラメータを設定する。

【0094】

また、図13は、第4の実施形態のディスク装置管理手段230であり、図3のミラー部235に替わり、236のコピー部を有する。231のエラー監視部は、ディスク装置のエラー発生状況を監視し、ディスク装置のエラー発生回数が「エラー回数規定値レベル1」を超えた場合にディスク装置からスペアディスク装置へのコピー開始を指示し、コピー終了後は、閉塞のステータスにする。237は、「エラー回数規定値レベル1」の値を再設定するエラー回数規定値変更部である。

以上が、本実施形態のアレイ型ディスク装置構成である。

【0095】

(2) ディスク装置2重障害予防動作

本実施形態は、ディスクアレイ(RAID)グループを形成する複数のディスク装置の故障ポテンシャルが高くなっている状態において、スペアディスク装置への予防コピー開始契機となるエラー発生回数の規定値を動的に変更することでディスク装置2重障害の確率を下げるものである。

【0096】

次に、ディスク装置2重障害予防動作を図14のフローチャートを用いて説明する。

【0097】

前提として、各ディスク装置301~307のエラー発生状況は、エラーカウンタ部232によってカウントし、ディスク装置管理テーブル240に継続的に設定されているものとする。図11は、ディスクアレイ(RAID)グループを構成するディスク装置301~305に対して、独立に実施するものとする。また、ディスクアレイ(RAID)グループを構成する「ディスク装置No.」の1と3は、エラー回数が多くなっており、潜在的にディスク装置の2重障害のポテンシャルが高くなっている状態を想定する。

【0098】

まず、エラー監視部231が、監視の対象としている該当ディスク装置のディスク装置管理テーブル240の「エラー回数カウンタ」の値が「エラー回数規定値レベル1」に達したかどうかを判定する(ステップ3001)。達した場合は、該当ディスク装置の内容をスペアディスク装置にコピーし、移行する処理を行う(ステップ3002)。達していない場合は、「エラー回数カウンタ」の値が「エラー回数補助規定値」に達したかどうかを判定する(ステップ3004)。達していない場合は、引き続きステップ3001を繰り返す。達していた場合は、同一ディスクアレイ(RAID)グループを構成する対象以外のディスク装置でエラー回数のカウンタの値が「エラー補助規定値」に達しているもの

があるかどうかを判定する(ステップ3005)。達しているものがない場合は、引き続きステップ3001を繰り返す。達しているものがある場合は、ディスクアレイ(RAID)グループを構成する全ディスク装置の「エラー回数規定値レベル1」の値を下げ(ステップ3006)、引き続きステップ3001を繰り返す。

【0099】

「エラー回数規定値レベル1」の値の再設定はエラー回数規定値変更部237で実施する。再設定する値は、「エラー回数規定値レベル1」と「エラー補助規定値」の中間値としておくなど任意の値でよい。また、ステップ3004、ステップ3005では、同一RAIDグループを構成する対象以外のディスク装置でエラー回数のカウンタの値が「エラー補助規定値」に達しているものがあるかどうかを判定基準としたが、たとえば、ディスクアレイ(RAID)グループを構成する全ディスク装置の「エラー回数カウンタ」の値の合計値を判定基準としてもよい。

以上が、ディスク装置2重障害予防動作である。

【0100】

(3) 効果

本実施形態によれば、スペアディスク装置に予防コピーしておくアレイ型ディスク装置において、ディスクアレイ(RAID)グループを形成する複数のディスク装置の故障ポテンシャルが高くなっている状態において、ディスク装置2重障害の確率を下げる信頼性の高いアレイ型ディスク装置を提供することができる。

なお、第4の実施形態は、エラー回数の判断基準となる規定値を動的に変更するものであり、第1、第2、第3の実施形態と組み合わせて適用してもよい。

【0101】

また、第1、第2の実施形態において、第3の実施形態のデータ復旧手段270を適用することで、ディスク装置障害を契機としたデータ復旧中における、あるディスク装置のセクタ読み出し障害にも対応することができる。

【0102】

(第5の実施形態)

次に、本発明の第5の実施形態を示す。第5の実施形態は、本発明が解決しようとする課題の第4の目的を達成するためのものである。

図15は本発明の第5の実施形態のアレイ型ディスク装置の構成を示す説明図である。本実施例のアレイ型ディスク装置は、ホストコンピュータ100とコマンド及びデータの入出力を行う複数のホストI/Fを具備した単一または複数のチャンネルコントローラ(1101)とホストコンピュータとの入出力データを一時的に蓄えておくキャッシュメモリ(1301)とホストコンピュータの入出力データを記憶しておくディスク装置(1601~1605)とディスク装置の制御を行うための単一または複数のディスク装置I/F1551を具備した単一または複数のディスクコントローラA(1401)と、同じく単一または複数のディスクドライブI/F1552を具備した単一または複数のディスクコントローラB(1402)がアクセス可能な共有メモリ1302と、チャンネルコントローラ(1101)ーキャッシュメモリ(1301)ー共有メモリ1302ーディスクコントローラA及びB(1401、1402)間のデータ転送及び通信を行うためのシステム・バス(1201、1202)より構成される。

また、ディスク装置D1(1601)、D2(1602)、D3(1603)、P(1604)は、ディスクアレイ(RAID)構成により冗長度を有している。

【0103】

ホストコンピュータ(100)からライトデータを受領したチャンネルコントローラ(1101)はキャッシュメモリ(1301)に退避すると共にディスクコントローラA(1401)またはディスクコントローラB(1402)に対して、キャッシュメモリ(1301)にあるライトデータをディスク装置(1601~1604)に書き込むよう指示する。またホストコンピュータ(100)からデータリード要求を受領したチャンネルコントローラはディスクコントローラA(1401)または、ディスクコントローラB(140

2) に対し、ディスク装置(ドライブ)(1601~1604)よりデータを読み出し、キャッシュメモリ(1301)に転送するよう指示する。指示を受けたディスクコントローラA(1401)またはディスクコントローラB(1402)はディスク装置(1601~1604)よりデータを読み出し、キャッシュメモリ(1301)に転送したのち、チャンネルコントローラ(1101)にデータ読み出し完了を報告する。報告を受けたチャンネルコントローラ(1101)データをキャッシュ(1301)よりホストコンピュータ(100)に転送する。

【0104】

図16はディスク装置D1(1601)でリードエラーが発生した場合の本発明の二重障害発生を防止したデータ回復を説明する図である。

【0105】

ディスク装置(ドライブ)D1(1601)上のデータD1_{D1}(2001)のリードエラーを検出したディスクコントローラA(1401)またはディスクコントローラB(1402)は共有メモリ(1302)上のディスク装置(ドライブ)情報(2101)を更新し、リードエラーとなったディスク装置D1(1601)のデータの冗長データをディスク装置D2(1602)のデータD2_{D1}(2002)、ディスク装置D3(1603)のデータデータD3_{D1}(2003)、ディスク装置DP(1604)のデータデータP_{D1}(2004)を読み出してキャッシュメモリ(1301)にそれぞれ、データD2_{D1}(2302)、データD3_{D1}(2303)、データP_{D1}(2304)として転送したのち、データD2_{D1}(2302)、データD3_{D1}(2303)、データP_{D1}(2304)を用いて冗長度計算により、ディスク装置D1(1601)のデータD1_{D1}(2301)を回復し、キャッシュメモリ(1301)に格納する。

【0106】

図17はディスク装置情報(2101)の構成要素を示す概略図である。

ディスク装置情報(2101)はデータリードエラーの発生回数を示す障害カウンタ(エラー発生回数カウンタ)(3001)、ディスク装置S(1605)へのデータ移行のコピー完了位置を示すコピーカウンタ(3002)、ディスク装置のリード/ライト可否情報等を示すディスク装置状態(3003)から構成され、障害カウンタ(エラー発生回数カウンタ)(3001)、コピーカウンタ(3002)の初期値は0、ディスク装置状態(3003)の初期値は“通常状態”である。

【0107】

図18はディスクドライブ状態(2101)のディスク装置状態が“通常状態”でディスク装置D1(1601)でデータリードエラーが発生した場合の状態変更処理を示すフローチャートである。

【0108】

ディスクコントローラA(1401)またはディスクコントローラB(1402)はディスク装置D1(1601)からのデータリードがエラーになった場合、ステップ4001で前述のように共有メモリ(1302)にあるディスク装置D1(1601)に関するディスク装置情報(2101)の障害カウンタ(3001)を加算する。次にステップ4002で障害カウンタ(2101)がしきい値N1を超えたかどうか判定する。もししきい値N1を超えた場合、ディスクコントローラA(1401)または、ディスクコントローラB(1402)はディスク装置D1(1601)が近い将来完全にデータ読み出しができなくなる可能性があると考え、ステップ3003でディスク装置情報(2101)のディスク装置状態(3003)を“データ移行中”に変更し、ステップ3003でディスク装置D1(1601)のデータD1_{D1}(2001)~D1_{Dm}(200n)をキャッシュメモリ(1301)上にデータD1_{D1}(2301)~D1_{Dm}(230n)として読み込み、順次ディスク装置S(1605)に書き込むことにより、ディスク装置D1(1601)のデータをディスクドライブS(1605)に移行する。またこの時ディスク装置S(1605)にデータD1_{Dm}(0 ≤ Dm ≤ Dn)移行毎にディスクドライブ情報(2101)のコピーカウンタをDmに更新する。

【0109】

図19はディスク装置状態(2101)のディスク装置状態が“データ移行中”でディスク装置D1(1601)においてデータリードエラーが発生した場合の状態変更処理を示すフローチャートである。

【0110】

ディスクコントローラA(1401)またはディスクコントローラB(1402)はディスク装置D1(1601)からのデータリードがエラーになった場合、ステップ5001で前述のように共有メモリ(1302)にあるディスク装置D1(1601)に関するディスク装置情報(2101)の障害カウンタ(エラー発生回数カウンタ)(3001)を加算する。次にステップ5002で障害カウンタ(エラー発生回数カウンタ)(2101)が規定値N2を超えたかどうか判定する。規定値N2を越えていた場合、ディスク装置状態を“警告”に変更し、ステップ5004でデータ移行元のデータD1_{D1}(2001)～D1_{Dm}(200n)をディスク装置D1(1601)から読み出すのではなく、ディスクアレイのレイド(RAID)機能を用いてディスク装置D2～P(1602～1604)から読み出して冗長度計算によって求める方式に変更する。

【0111】

図20はディスク装置情報(2101)のディスク装置状態(3003)が“通常状態”または“データ移行中”時におけるディスク装置D1(1601)のデータD1_{Dm}($0 \leq D_m \leq D_n$)のリード方式を示すフローチャートである。

【0112】

ステップ6001でディスク装置D1(1601)からデータD1_{Dm}をリードしてキャッシュに転送する。ステップ6002でリードエラーがどうか判定し、リードエラーが発生した場合、ステップ6003で前述の通り冗長度を持つアレイディスクグループを構成するディスク装置D2(1602)、ディスク装置D3(1603)、ディスク装置P(1604)を用いて、ディスク装置D1(1601)のデータD1_{Dm}を作成する。

【0113】

また、ディスク装置情報(2101)のディスク装置状態(3003)が“通常状態”または“データ移行中”時におけるディスク装置D1(1601)のデータD1_{Dm}($0 \leq D_m \leq D_n$)のライト方式は更新ライトデータをD1_{D1}(2301)とした場合、ディスクコントローラA(1401)またはディスクコントローラB(1402)はディスク装置D1(1601)の当該ブロック位置に存在しているデータD1_{D1}(2001)を読み出し、キャッシュメモリ(1301)上に旧データO1_{D1}(2311)として格納する。次にディスク装置P(1604)よりデータP_{D1}(2004)を読み出し、キャッシュメモリ(1301)上に旧パリティデータP_{OD1}(2314)として格納する。次にディスクコントローラA(1401)またはディスクコントローラB(1402)は更新データD1_{D1}(2301)、旧データO1_{D1}(2311)及び旧パリティデータP_{OD1}(2314)を用いて、排他的論理和演算により、新パリティデータP_{D1}(2304)を生成し、キャッシュメモリ(1301)に格納する。次にディスクコントローラA(1401)またはディスクコントローラB(1402)は更新データD1_{D1}(2301)をディスク装置D1(1601)及びディスク装置S(1605)に書き込むとともに、先に生成した新パリティデータP_{D1}(2304)をディスク装置P(2004)に書き込む。

【0114】

図21はディスク装置情報(2101)のディスク装置状態(3003)が“警告”時におけるディスク装置D1(1601)のデータD1_{Dm}($0 \leq D_m \leq D_n$)のリード方式を示すフローチャートである。

【0115】

ステップ7001で前述の通り冗長度を持つアレイディスクグループを構成するディスク装置D2(1602)、ディスク装置D3(1603)、ディスク装置P(1604)を用いて、ディスク装置D1(1601)のデータD1_{Dm}を作成する。ステップ7002でディスク装置D2～P(1602～1604)のいずれかのディスク装置でデータD

x_{D_m} (DX:D2orD3orP) リードエラーが発生した場合、ステップ 7003 で D_m とディスク装置情報 (2001) のコピーカウンタ (3002) と比較し、コピーカウンタ以下であれば既にディスク装置 S (1605) に当該データの移行は完了しているため、ステップ 7004 でディスク装置 S よりデータ $D1_{D_m}$ をリードする。またコピーカウンタを超える場合は、ステップ 7005 でディスク装置 D1 (1601) より $D1_{D_m}$ をリードする。またこの時 $D1_{D_m}$ を用いてリードエラーとなった Dx_{D_m} を回復してもよい、例えば $D2_{D_m}$ がリードエラーとなった場合、冗長度を持つアレイディスクグループを構成する $D1_{D_m}$ と $D3_{D_m}$ と P_{D_m} より $D2_{D_m}$ を回復し、ディスク装置 D2 (1602) に交替媒体領域を設定し、その領域に回復した $D2_{D_m}$ を書き込んでもよい。

【0116】

またディスクドライブ情報 (2101) のディスク装置状態 (3003) が“警告”時におけるディスク装置 D1 (1601) のデータ $D1_{D_m}$ ($0 \leq D_m \leq D_n$) のライト方式はライトデータを $D1_{D1}$ (2301) とした場合、ディスクコントローラ A (1401) またはディスクコントローラ B (1402) はディスク装置 D2 (1602)、及びディスク装置 D3 (1603) の当該ブロックにて冗長度を有しているデータ $D2_{D1}$ (2002)、及びデータ $D3_{D1}$ (2003) を読み出し、キャッシュメモリ (1301) 上にそれぞれ旧データ $O2_{D1}$ (2312) 及び旧データ $O3_{D1}$ (2313) として格納する。次にディスクコントローラ A (1401) またはディスクコントローラ B (1402) は更新データ $D1_{D1}$ (2301)、旧データ $O2_{D1}$ (2312)、旧データ $O3_{D1}$ (2313) を用いて、排他的論理和演算により、新パリティデータ P_{D1} (2304) を生成し、キャッシュメモリ (1301) に格納する。次にディスクコントローラ A (1401) またはディスクコントローラ B (1402) は更新データ $D1_{D1}$ (2301) をディスク装置 D1 (1601) 及びディスク装置 S (1605) に書き込むとともに、先に生成した新パリティデータ P_{D1} (2304) をディスク装置 P (2004) に書き込む。

【図面の簡単な説明】

【0117】

【図1】 本発明の第1の実施形態における装置構成図である。

【図2】 本発明の第1の実施形態におけるディスク装置管理テーブルの説明図である。

。

【図3】 本発明の第1の実施形態におけるディスク装置管理手段の説明図である。

【図4】 本発明の第1の実施形態における予防スペアコピー動作のフローチャートである。

【図5】 本発明の第2の実施形態におけるディスク装置管理テーブルの説明図である。

。

【図6-1】 本発明の第2の実施形態におけるダイナミックミラーリング動作のフローチャートである。

【図6-2】 本発明の第2の実施形態におけるダイナミックミラーリング動作のフローチャートである。

【図7】 本発明の第3の実施形態における装置構成図である。

【図8】 本発明の第3の実施形態におけるディスク装置管理テーブルの説明図である。

。

【図9】 本発明の第3の実施形態におけるディスク装置管理部の説明図である。

【図10】 本発明の第3の実施形態におけるセクタ障害復旧動作のフローチャートである。

【図11】 本発明の第3の実施形態におけるセクタ障害復旧動作時のライト動作のフローチャートである。

【図12】 本発明の第4の実施形態におけるディスク装置管理テーブルの説明図である。

【図13】 本発明の第4の実施形態におけるディスク装置管理手段の説明図である。

【図14】 本発明の第4の実施形態におけるディスク装置2重障害予防動作のフロー

チャートである。

【図 1 5】 本発明の第 5 の実施形態における装置構成を示す図である。

【図 1 6】 本発明の第 5 の実施形態における装置構成の詳細を示す図である。

【図 1 7】 本発明の第 5 の実施形態における装置構成の部分詳細を示す図である。

【図 1 8】 本発明の第 5 の実施形態における動作フローチャートを示す図である。

【図 1 9】 本発明の第 5 の実施形態における別の動作フローチャートを示す図である。

。

【図 2 0】 本発明の第 5 の実施形態における他の動作フローチャートを示す図である。

。

【図 2 1】 本発明の第 5 の実施形態におけるさらに別の本発明の原理を示す動作フローチャートを示す図である。

【符号の説明】

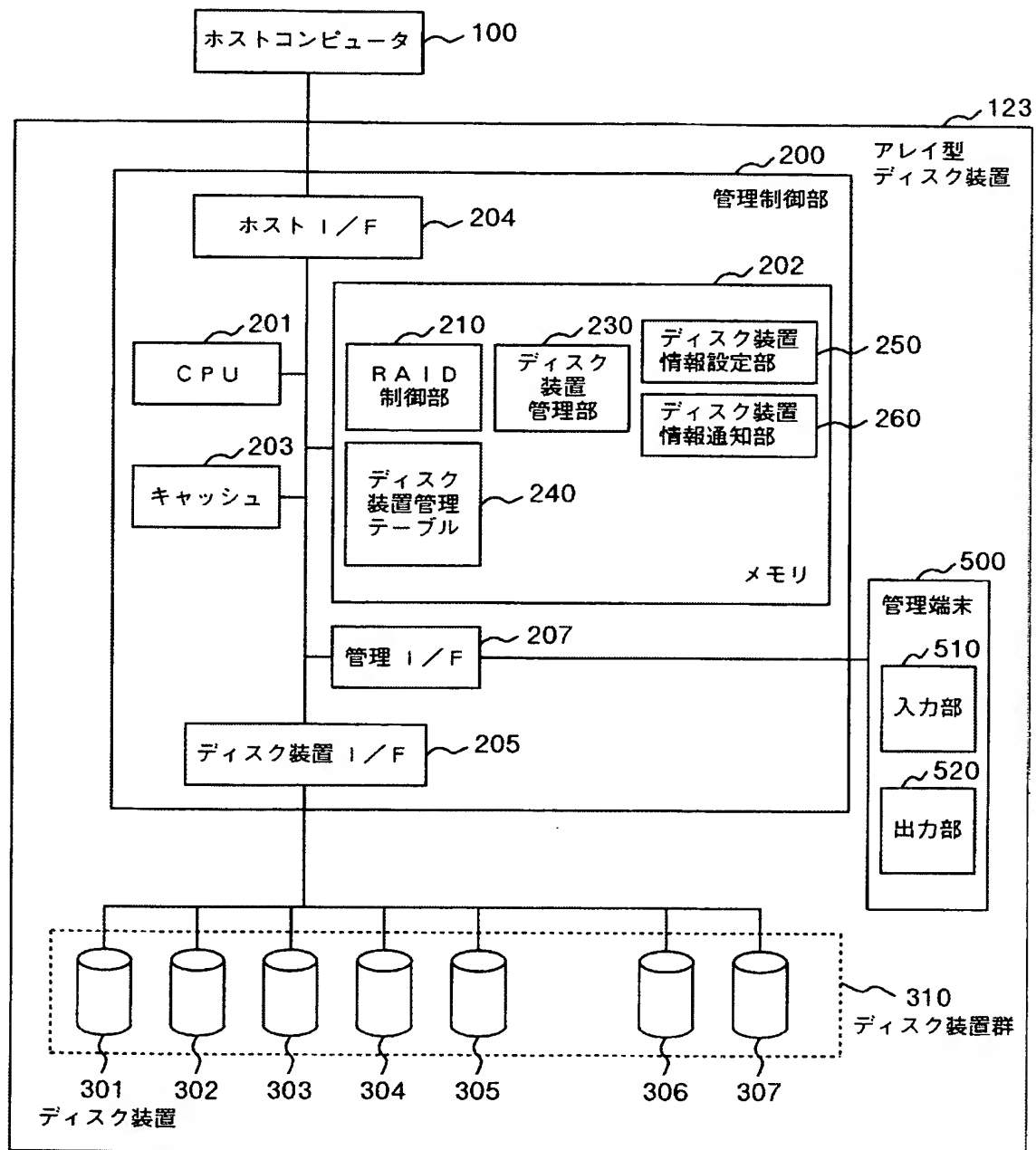
【0 1 1 8】

1 0 0…ホストコンピュータ、1 2 3…アレイ型ディスク装置、2 0 0…アレイ型ディスク装置の管理制御部、3 1 0…ディスク装置群、5 0 0…管理端末、
2 0 1…CPU、2 0 2…メモリ、2 0 3…キャッシュ、2 0 4…ホスト I/F、2 0 5…ディスク装置 I/F、2 0 7…管理 I/F であり、それぞれを結線する。2 1 0…アレイ型ディスク装置制御する RAID 制御部、2 3 0…ディスク装置管理部、2 4 0…ディスク装置管理テーブル、2 5 0…ディスク装置情報設定部、2 6 0…ディスク装置情報通知部、ディスク装置 3 0 1～3 0 7…ディスク装置
5 1 0…入力部、5 2 0…出力部。

【書類名】 図面
【図 1】

図 1

第 1 の実施形態の装置構成図



【図 2】

図 2

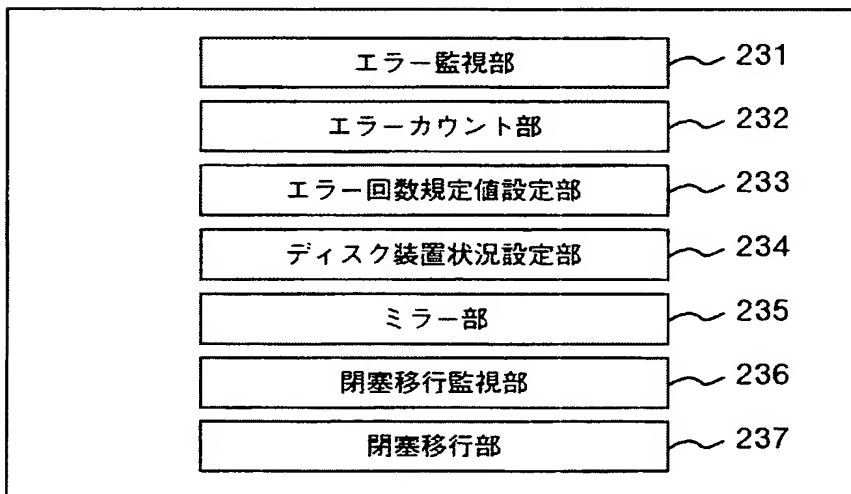
ディスク装置管理テーブル

ディスク 装置 No.	エラー回数 カウンタ	エラー回数 規定値 レベル 1	エラー回数 規定値 レベル 2	スベア ビット	ディスク 装置状況	ペア ディスク 装置
0	5	5 0	9 0	No	正常	—
1	5	5 0	9 0	No	正常	—
2	5	5 0	9 0	No	正常	—
3	5	5 0	9 0	No	正常	—
4	6 0	5 0	9 0	No	ミラー	No. 5
5	5	5 0	9 0	Yes	ミラー	No. 4
6	5	5 0	9 0	Yes	正常	—
...						

【図 3】

図 3

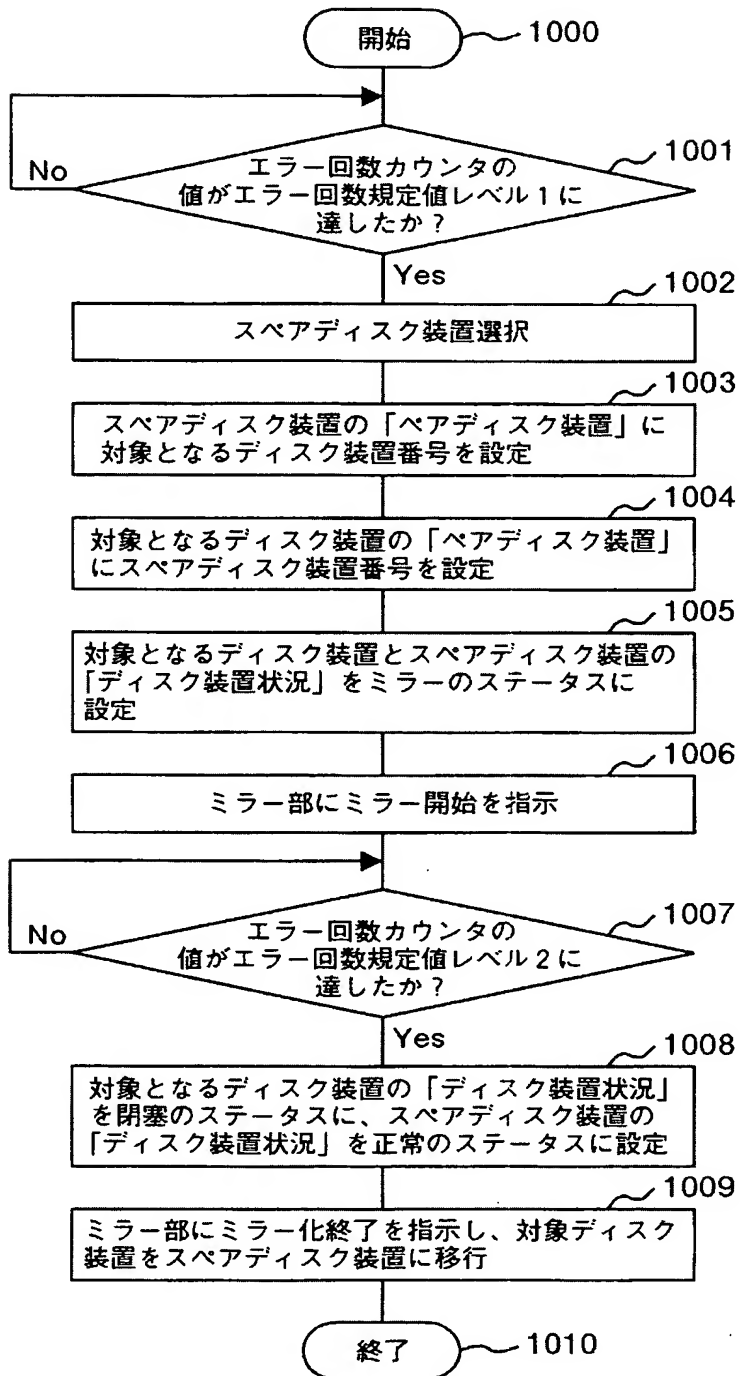
ディスク装置管理部



【図 4】

図 4

予防スベアコピー動作



【図 5】

図 5

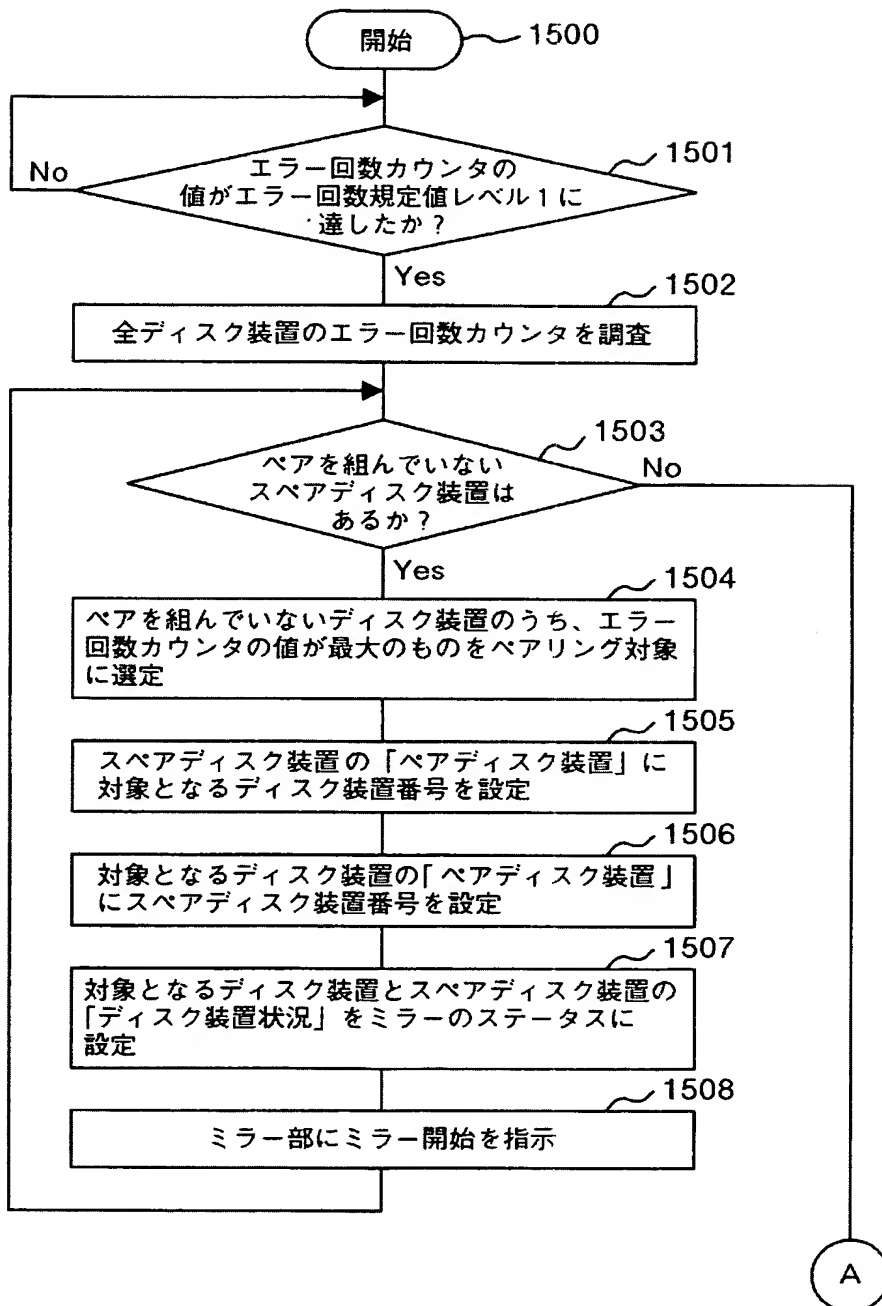
ディスク装置管理テーブル

ディスク 装置 No.	エラー回数 カウンタ	エラー回数 規定値 レベル 1	エラー回数 規定値 レベル 2	スベア ビット	ディスク 装置状況	ペア ディスク 装置
0	1 0	3 0	9 0	No	正常	—
1	5	3 0	9 0	No	正常	—
2	3 5	3 0	9 0	No	ミラー	No. 6
3	5	3 0	9 0	No	正常	—
4	2 0	3 0	9 0	No	ミラー	No. 5
5	5	3 0	9 0	Yes	ミラー	No. 4
6	5	3 0	9 0	Yes	ミラー	No. 2
...						

【図 6-1】

図 6-1

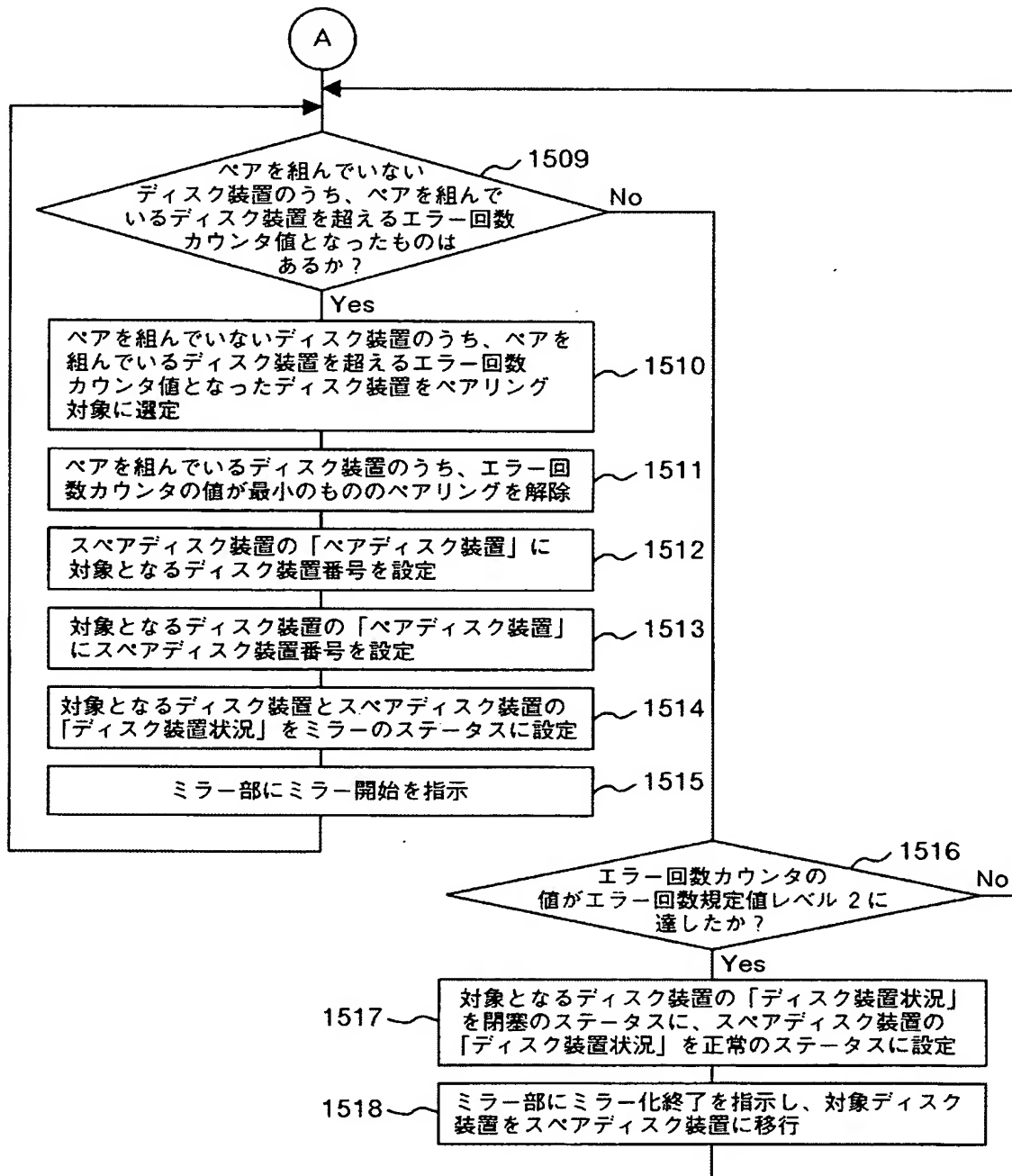
ダイナミックミラーリング動作



【図 6-2】

図 6-2

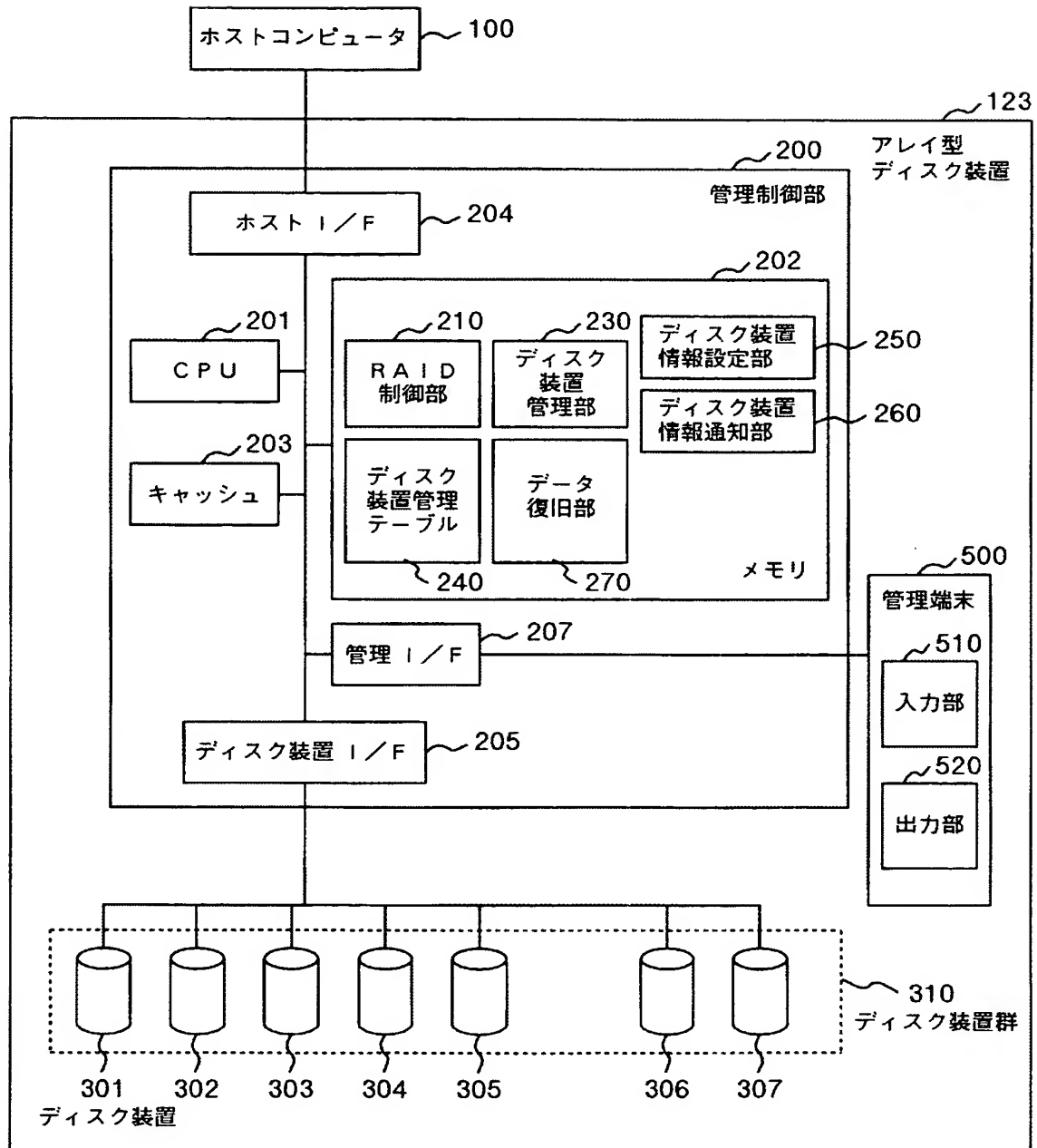
ダイナミックミラーリング動作



【図 7】

図 7

第3の実施形態の装置構成図



【図 8】

図 8

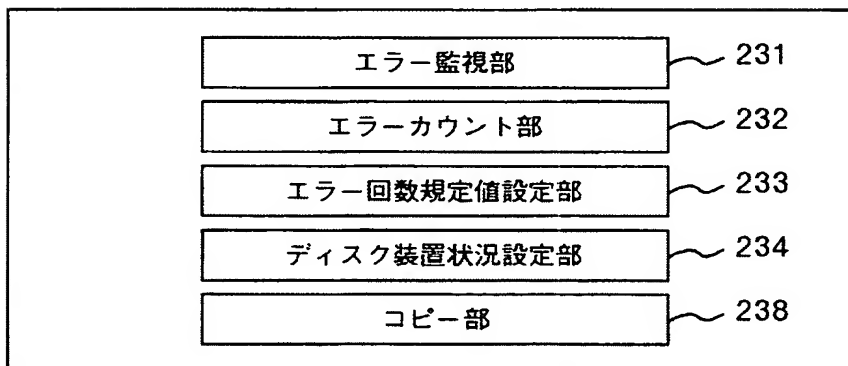
ディスク装置管理テーブル

ディスク 装置 No.	エラー回数 カウンタ	エラー回数 規定値 レベル 1	スベア ビット	ディスク 装置状況	ペア ディスク 装置
0	5	80	No	復旧中	—
1	5	80	No	復旧中	—
2	5	80	No	復旧中	—
3	5	80	No	復旧中	—
4	85	80	No	仮閉塞	No. 5
5	5	80	Yes	復旧中	No. 4
6	5	80	Yes	正常	—
...					

【図 9】

図 9

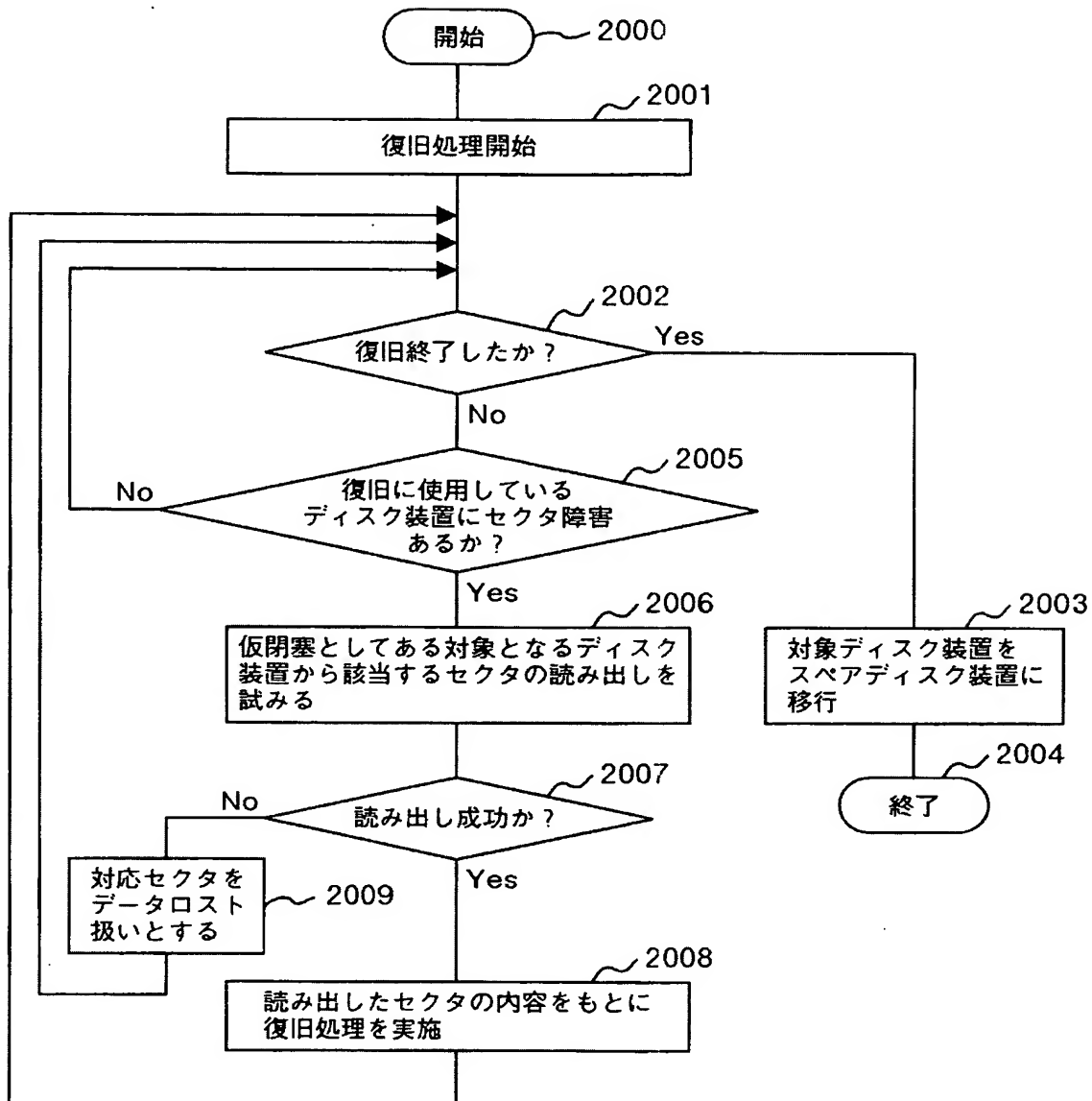
ディスク装置管理部



【図 10】

図 10

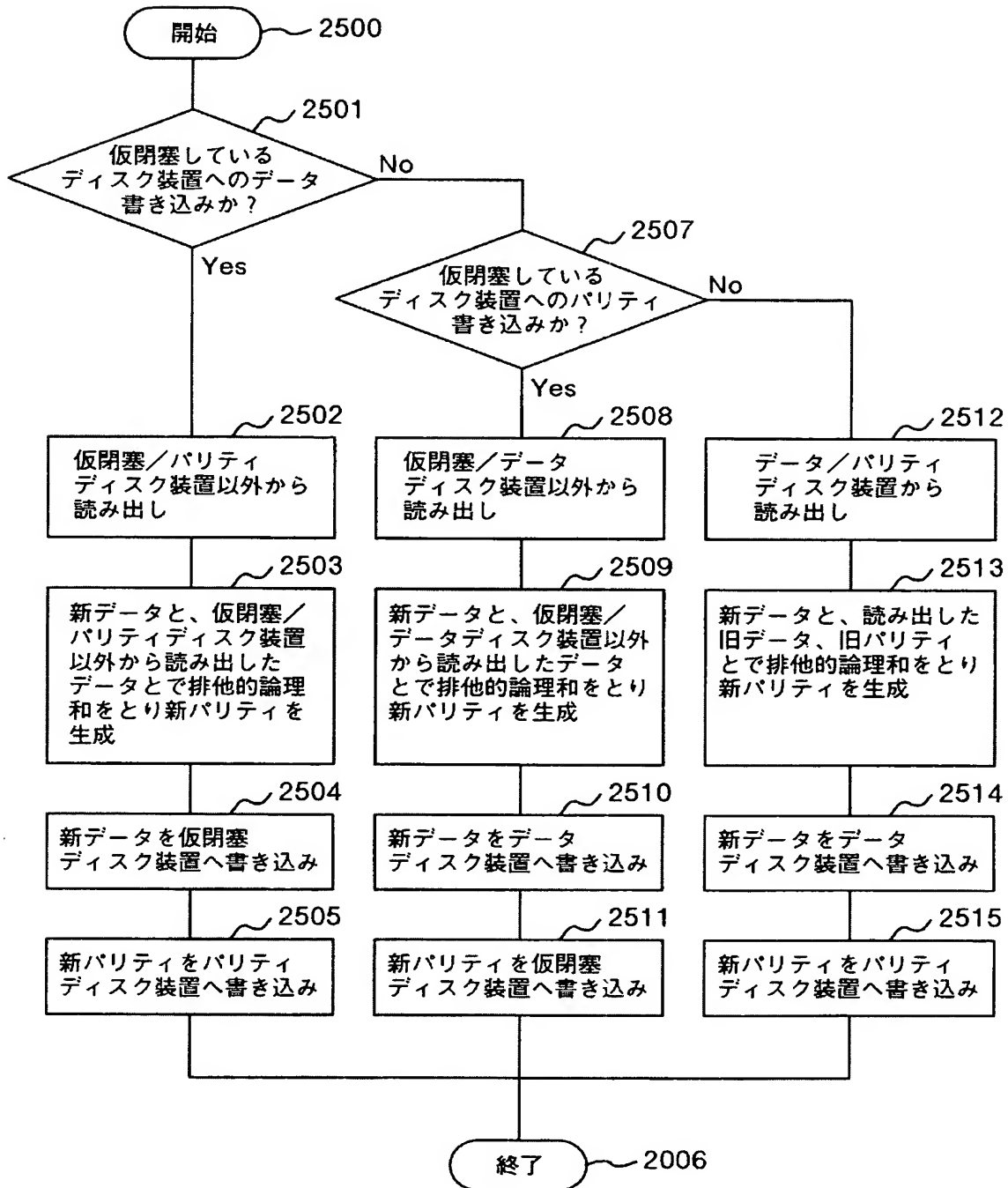
セクタ障害復旧動作



【図 11】

図 11

セクタ障害復旧動作時のライト動作



【図 12】

図 12

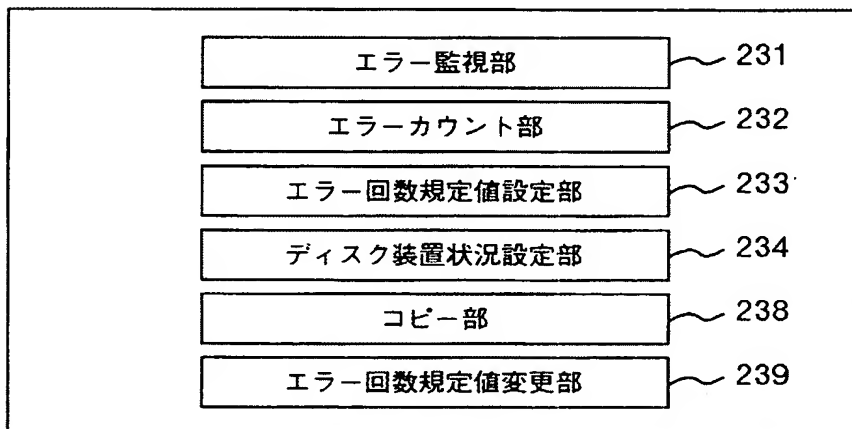
ディスク装置管理テーブル

ディスク 装置 No.	エラー回数 カウンタ	エラー回数 規定値 レベル1	エラー回数 補助規定値	スベア ビット	ディスク 装置状況	ペア ディスク 装置
0	5	60	40	No	正常	—
1	40	60	40	No	正常	—
2	5	60	40	No	正常	—
3	40	60	40	No	正常	—
4	5	60	40	No	正常	—
5	5	80	40	Yes	正常	—
6	5	80	40	Yes	正常	—
...						

【図 13】

図 13

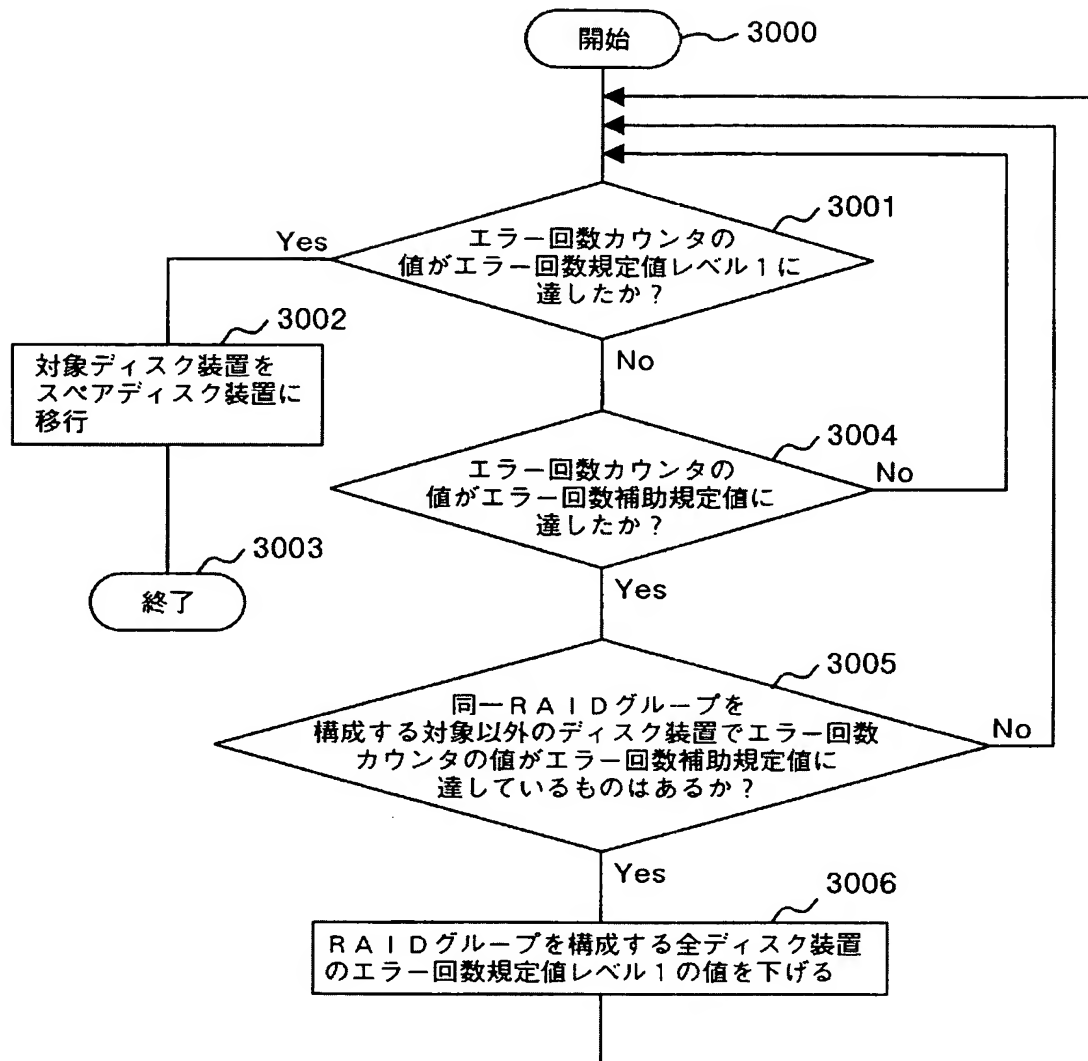
ディスク装置管理部



【図 14】

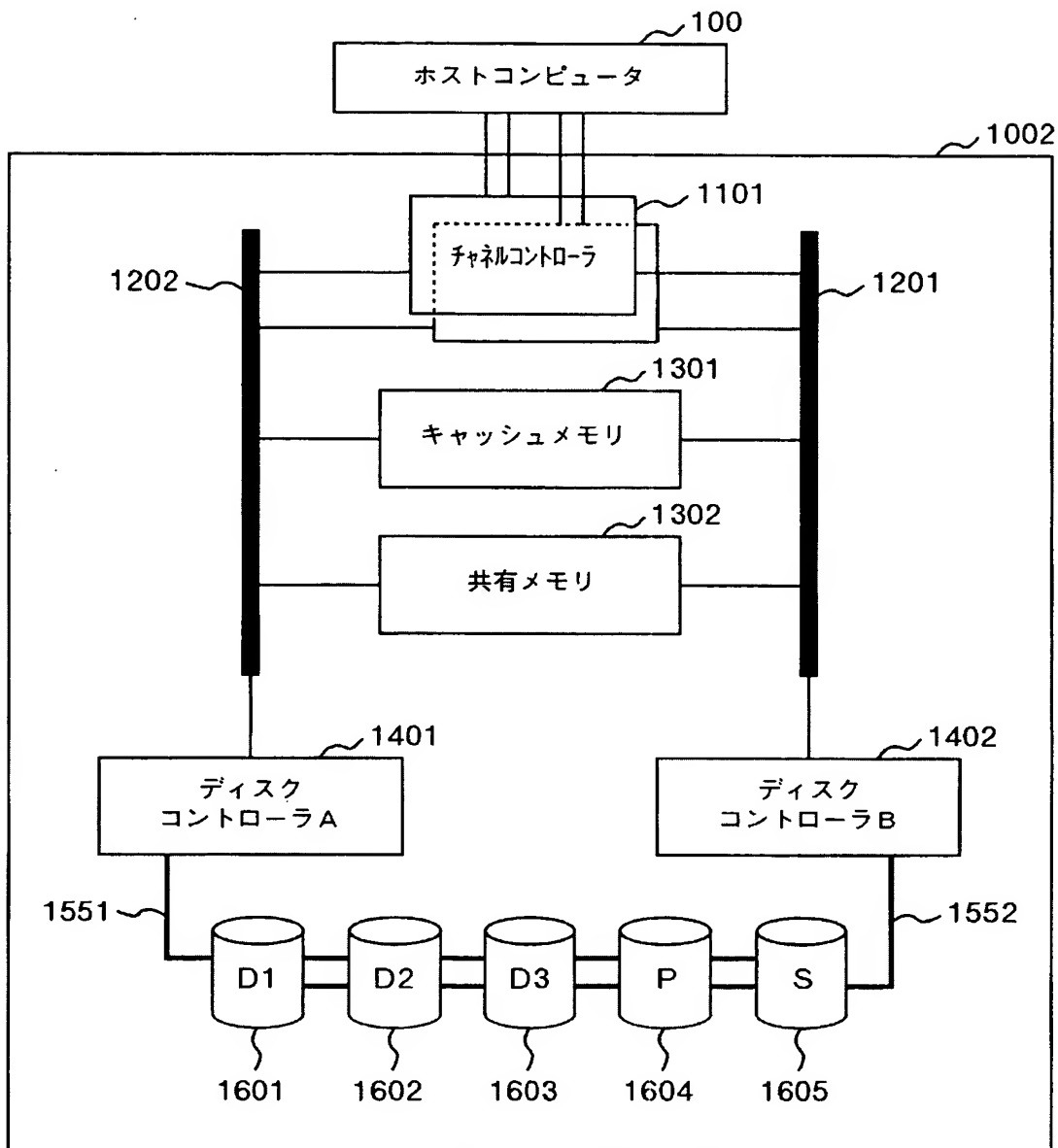
図 14

ディスク装置 2 重障害予防動作



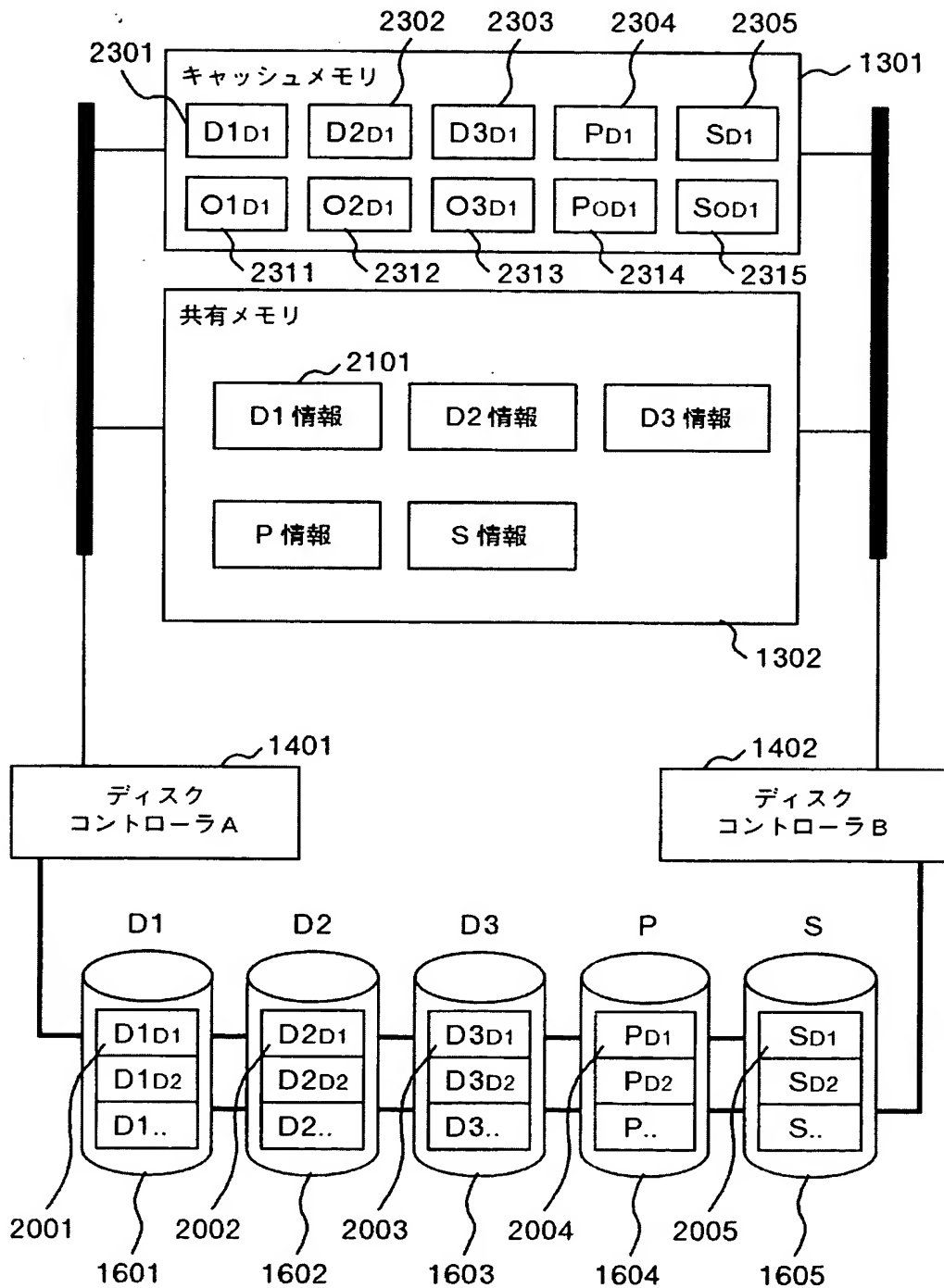
【図 15】

図 15



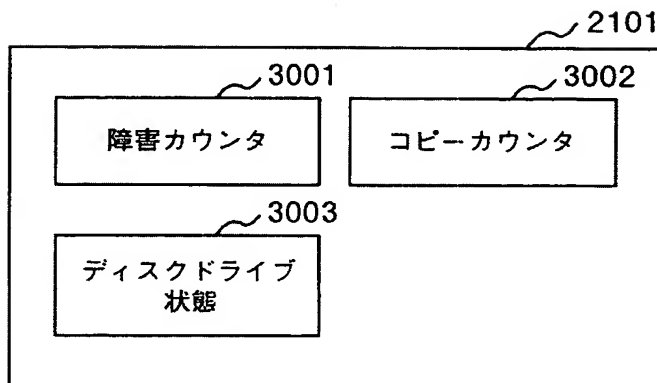
【図 16】

図 16



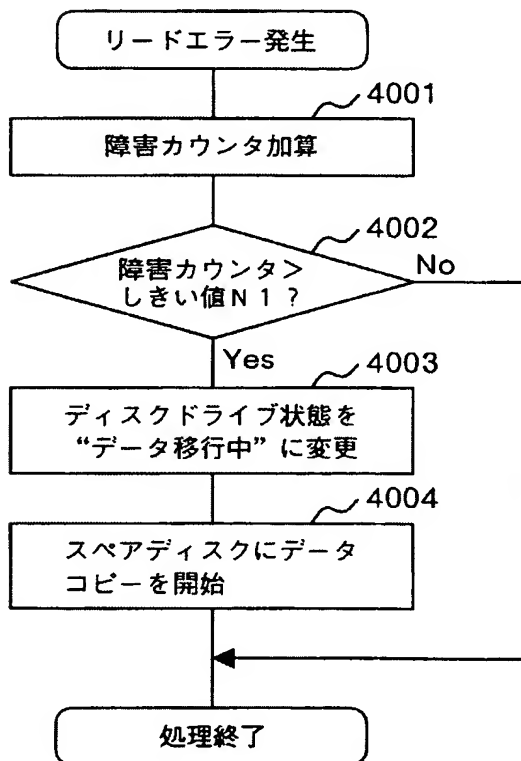
【図 17】

図 17



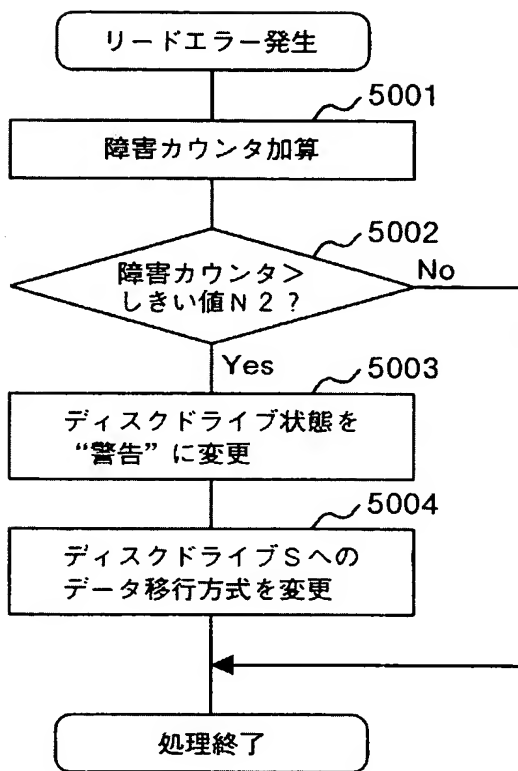
【図 18】

図 18



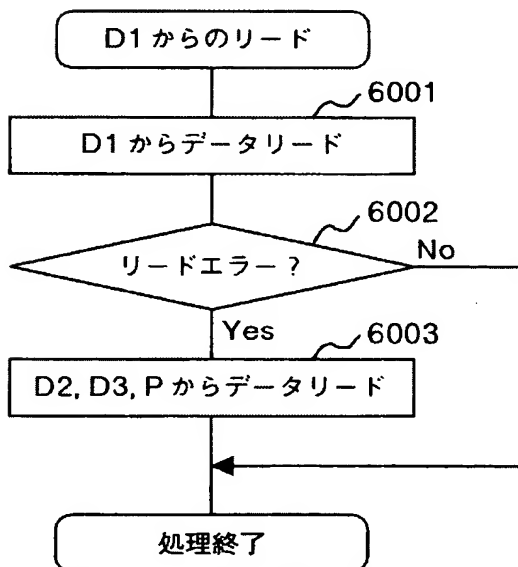
【図 19】

図 19



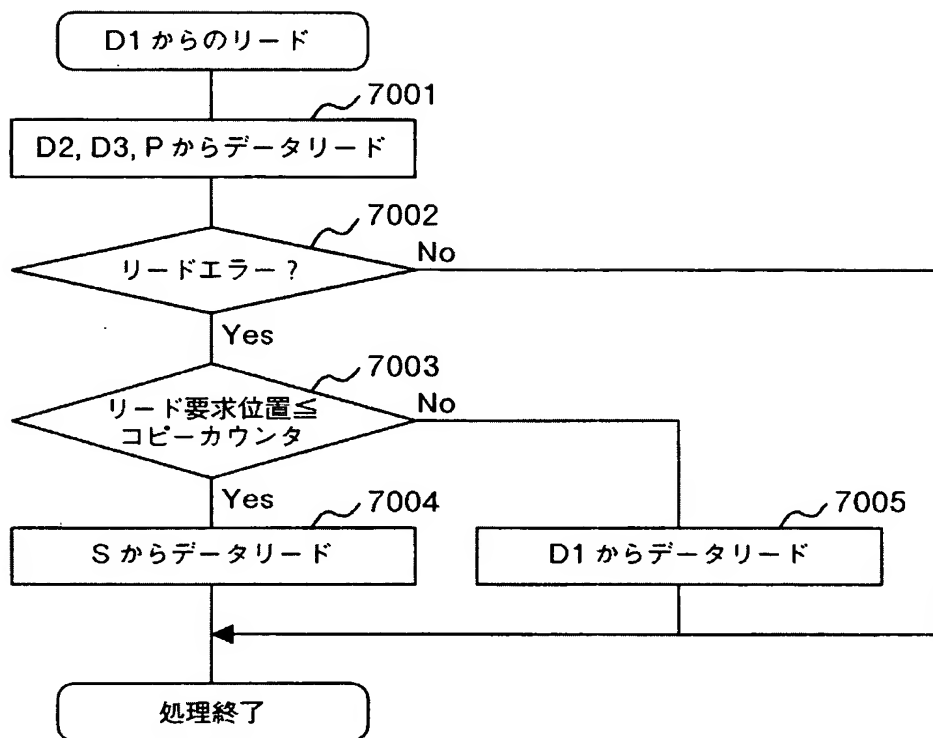
【図 20】

図 20



【図 21】

図 21



【書類名】 要約書**【要約】**

【課題】 アレイ型ディスク装置において、冗長なディスク装置を1台分備えたディスクアレイグループ(RAID)構成の場合、1台のディスク装置が故障している状態でもう一台のディスク装置からの読み出しができなくなるディスク2重障害を防止する。

また、かかる冗長性のあるアレイディスクグループ構成において、ディスク装置間のデータ移行に際し、ホストレスポンスを低下されることなく、かつデータを失うことなく、データ移行を完了する手段を提供する。

【解決手段】 アレイ型ディスク装置において、ディスク装置のエラー発生状況を監視し、エラー発生回数が規定値を超えた場合にディスク装置とスペアディスク装置とのミラーリング開始を指示するエラー監視手段と、ディスク装置とスペアディスク装置とのミラーリングを行うミラー手段とを設ける。また、アレイ型ディスク装置において、ディスク装置のエラー発生状況を監視し、エラー発生回数が規定値を超えた場合にディスク装置のステータスが仮閉塞状態となるよう指示するエラー監視手段と、データ復元中にディスクアレイを構成する他のディスク装置からの読み出しができない場合、仮閉塞状態のディスク装置からの読み出しを行うことでデータ復元を行うデータ復旧手段とを設ける。

さらに、ディスク装置D1(1601)からディスク装置S(1605)にデータを移行する際、ディスク装置D1(1601)で発生したリードエラー回数を示す障害カウンタを共有メモリ(1302)に記憶し、読み出し障害カウントに応じて、移行元のデータをディスク装置D1(1601)から読み込むか、あるいは、アレイディスクグループを構成するディスク装置D2(1602)、ディスク装置D3(1603)、ディスク装置P(1604)から読み込み、冗長度計算によって復旧するか選択可能とする。

【選択図】 図1

特願 2 0 0 3 - 3 3 5 4 6 5

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所